

U.S. Application No. 10/520,224
Attorney Docket: 043043-0358757

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | |
|--|----------------------------|
| In re PATENT APPLICATION OF | Confirmation No.: 2578 |
| Mueller-Hermelink et al. | Group Art Unit: 1642 |
| Application No. 10/520,224 | Examiner: Reddig, Peter J. |
| Filed: August 19, 2005 | |
| Title: Neoplasm Specific Antibodies and Uses Thereof | |

DECLARATION UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

We, the undersigned, do hereby declare and state as follows:

1. We are the inventors of the claimed subject matter of the above-identified U.S. patent application.

2. Brandlein *et al.* (Amer. Assoc. Cancer Res. 43:970 abstract #4803 (2002)) and Brandlein *et al.* (Human Antibodies 11:107 (2002)), hereinafter, the Brandlein references were cited in rejections under 35 U.S.C. §§ 102(a) and 103(a) of the Office Communication as allegedly anticipating or making obvious the claimed antibodies and antigen binding fragments. To the extent, if any, that the Brandlein references describe or suggest the subject matter claimed in the above-identified U.S. patent application, the subject matter described or suggested in the Brandlein references was derived from us, and was invented by us, rather than the other authors of the Brandlein references.

3. We further declare that all statements made herein of our own knowledge are true, that all statements made on information and belief are believed to be true, and that these statements were made with knowledge that willful false statements and the like so made are

U.S. Application No. 10/520,224
Attorney Docket: 043043-0358757

punishable by fine or imprisonment, or both (18 U.S.C. § 1001), and may jeopardize the validity of the application or any patent issued thereon.

30.9.08

Date

Hans-Konrad Mueller-Hermelink

Hans-Konrad Mueller-Hermelink

30.9.08

Date

Heinz Peter Vollmers

Heinz Peter Vollmers

Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bacterial secretion but not the affinity

Sergey M.Kipriyanov, Gerhard Moldenhauer¹,
Andrew C.R.Martin², Olga A.Kupriyanova and
Melvyn Little³

Recombinant Antibody Research Group (0445), Diagnostics and Experimental Therapy Programme, ¹Department of Molecular Immunology (0740), Tumor Immunology Programme, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany and ²Biomolecular Structure and Modeling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

³To whom correspondence should be addressed

Recombinant antibody fragments directed against cell surface antigens have facilitated the development of novel therapeutic agents. As a first step in the creation of cytotoxic immunoconjugates, we constructed a single-chain Fv fragment derived from the murine hybridoma OKT3, that recognizes an epitope on the ϵ -subunit of the human CD3 complex. Two amino acid residues were identified that are critical for the high level production of this scFv in *Escherichia coli*. First, the substitution of glutamic acid encoded by a PCR primer at position 6 of V_H framework 1 by glutamine led to a more than a 30-fold increase in the production of soluble scFv. Second, the substitution of cysteine by a serine in the middle of CDR-H3 additionally doubled the yield of soluble antibody fragment without any adverse effect on its affinity for the CD3 antigen. The double mutant scFv (Q,S) proved to be very stable *in vitro*: no loss of activity was observed after storage for 1 month at 4°C, while the activity of scFv containing a cysteine residue in CDR-H3 decreased by more than half. The results of production yield, affinity, stability measurements and analysis of three-dimensional models of the structure suggest that the sixth amino acid influences the correct folding of the V_H domain, presumably by affecting a folding intermediate, but has no effect on antigen binding.

Keywords: affinity/anti-human CD3/bacterial expression/single-chain Fv/solubility

Introduction

In recent years, the use of genetic engineering techniques has stimulated the development of antibody-like molecules for therapeutic and diagnostic uses (Winter and Milstein, 1991). Unlike glycosylated whole antibodies, fragments such as Fab and Fv can be easily produced in bacterial cells as functional antigen binding molecules (Better *et al.*, 1988; Skerra and Plückthun, 1988). To stabilize the association of the recombinant V_H and V_L domains, they have been linked in a single-chain Fv (scFv) construct with a short peptide that connects the carboxy terminus of one domain and the amino terminus of the other (Bird *et al.*, 1988; Huston *et al.*, 1988). In comparison with the much larger Fab, F(ab)₂ and IgG forms of monoclonal antibody from which they are derived, scFvs have more rapid blood clearance and better tumor penetration

(Milenic *et al.*, 1991; Yokota *et al.*, 1992; Adams *et al.*, 1993). ScFvs therefore represent potentially highly useful molecules for the targeted delivery of drugs, toxins or radionuclides to a tumor site.

The efficient expression of active antibody fragments in bacteria is clearly of great technological importance. However, as with the expression of some other heterologous proteins in *Escherichia coli*, the yield of functional product for some antibody fragments can be very low. Sometimes, PCR primer-induced errors can lead to the expression of non-reactive antibody fragments (McCartney *et al.*, 1995). Poor expression may also arise from differences in the translation machinery and folding pathways of eukaryotic and bacterial cells. For example, some nucleotide sequences encoding antibody variable regions were expressed as functional proteins in eukaryotic host cells but were unable to express a product in bacteria (Duenas *et al.*, 1995). Limiting factors for the efficient production of secreted antibody fragments in *E.coli* appear to be translocation to the periplasm (Ayala *et al.*, 1995) and folding in the periplasmic space (Knappik and Plückthun, 1995).

OKT3 is a murine monoclonal antibody (mAb) that recognizes an epitope on the ϵ -subunit of the human CD3 complex (Kung *et al.*, 1979; Van Wauwe *et al.*, 1980; Transy *et al.*, 1989). It has significant clinical utility. OKT3 has been widely used to suppress T cells and thereby prevent the rejection of transplants (Thistlethwaite *et al.*, 1984; Woodle *et al.*, 1991). Conversely, T cell activation and proliferation induced by OKT3 have been exploited to expand effector cells *ex vivo* for adoptive cancer immunotherapy (Yannelly *et al.*, 1990). As well as being used alone, the OKT3 mAb has been used as a component of bispecific antibodies to retarget cytotoxic T lymphocytes against tumor cells (Nitta *et al.*, 1990; Bohlen *et al.*, 1993) or virus infected cells (Sanna *et al.*, 1995). Recently, humanized versions of the OKT3 mAb have been expressed in COS cells (Woodle *et al.*, 1992; Adair *et al.*, 1994).

In this paper, we present the first example of the expression of an OKT3 derived scFv in *E.coli*. As part of the anti-CD3 scFv construction process, the PCR amplified OKT3 V_H gene was modified to improve its *in vivo* folding. Here we analyze the effect of two amino acid residues in the variable heavy chain domain on the yield, affinity and stability *in vitro* of anti-CD3 scFv.

Materials and methods

E.coli strains, plasmids and cell lines

E.coli K12 strain XL1-Blue (Stratagene, La Jolla, CA) was used as the cloning and expression host. For cloning, sequencing hybridoma-derived immunoglobulin variable regions and site-specific mutagenesis, pCR-Script SK(+) (Stratagene) was used. The scFv gene was assembled and expressed either in the plasmid pOPE51 (Kipriyanov *et al.*, 1994) or in pHOG21 (Kipriyanov *et al.*, 1996b). The hybridoma OKT3 producing a monoclonal antibody (IgG2a) against the CD3 human T cell antigen has been described previously (Kung *et al.*, 1979; Van

Wauwe *et al.*, 1980). The human CD3-positive acute T cell leukemia cell line Jurkat and a CD3-negative B cell line JOK-1 were used for flow cytometry.

Cloning of the variable regions

Isolation of mRNA from freshly subcloned hybridoma OKT3 cells and cDNA synthesis were performed as previously described (Dübel *et al.*, 1994). DNA coding for the light chain variable domain was amplified by PCR using the primers Bi5 and Bi8 that hybridize to the amino terminal portion of the κ chain constant domain and the framework 1 (FR1) region of the κ chain variable domain (Dübel *et al.*, 1994). For the amplification of DNA coding for the heavy chain variable domain, the primer Bi4 that hybridizes to the amino terminal portion of the γ chain constant 1 domain (Dübel *et al.*, 1994) and Bi3f that hybridizes to the FR1 region of the heavy chain (Gotter *et al.*, 1995; Kipriyanov *et al.*, 1996b) were used. The 50 μ l reaction mixture contained 10 pmol of each primer and 50 ng of hybridoma cDNA, 100 μ M each of dNTP, 1 \times Vent-buffer (Boehringer Mannheim, Mannheim, Germany), 5 μ g BSA and 1 U Vent DNA polymerase. 30 cycles of 1 min at 95°C, 1 min at 55°C and 2 min at 75°C were carried out in a thermocycler. The amplified DNA was purified with a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and blunt end ligated into an *Srf*I digested pCR-Script SK(+) (Stratagene) for dideoxy sequencing (Sanger *et al.*, 1977) and site-specific mutagenesis.

Construction of plasmids encoding scFv

The linker used in this study was a 17 amino acid tag-linker that includes a tubulin epitope recognized by mAb YOL1/34 (Breitling *et al.*, 1991). DNA coding for the variable domains of OKT3 was inserted into pOPE51 (Kipriyanov *et al.*, 1994) in two cloning steps using *Nco*I/*Hind*III for the heavy-chain DNA and *Eco*RV/*Bam*HI for the light-chain DNA. The whole scFv gene was recloned in pHOG21 (Kipriyanov *et al.*, 1996b) as a *Nco*I/*Bam*HI DNA fragment.

Construction of anti-CD3 mutants

Mutations were generated in the V_H domain derived from OKT3 by site-specific mutagenesis according to Kunkel *et al.* (1987). The amino acid substitution of Cys at position H100A by Ser and of Glu at position H6 by Gln was achieved using either primer SK1 5'-GTAGTCAAGGCTGTAATGATCATC or SK2 5'-GCCCCAGACTGCTGCAGCTGCAC or both.

E.coli expression and purification of scFv fragments

XL1-Blue *E.coli* cells (Stratagene) transformed with the scFv expression plasmid pHOG21 were grown overnight in 2 \times YT medium with 50 μ g/ml ampicillin and 100 mM glucose (2 \times YT_{GA}) at 37°C. Dilutions (1:50) of the overnight cultures in 2 \times YT_{GA} were grown as flask cultures at 37°C with shaking at 200 r.p.m. When cultures reached OD₆₀₀ = 0.8, bacteria were pelleted by centrifugation at 1500 g for 10 min and 20°C and resuspended in the same volume of fresh 2 \times YT medium containing 50 μ g/ml ampicillin and 0.4 M sucrose. IPTG was added to a final concentration of 0.1 mM and growth was continued at room temperature (20–22°C) for 20 h. The cells were harvested by centrifugation at 5000 g for 10 min and 4°C. The culture supernatant was retained and kept on ice. To isolate soluble periplasmic proteins, the pelleted bacteria were resuspended in 5% of the initial volume of ice-cold 50 mM Tris-HCl, 20% sucrose, 1 mM EDTA, pH 8.0. After a 1 h incubation on ice with occasional stirring, the spheroplasts were centrifuged at 30 000 g for 30 min and 4°C leaving the

soluble periplasmic extract as the supernatant and spheroplasts plus the insoluble periplasmic material as the pellet. The culture supernatant and the soluble periplasmic extract were combined, clarified by additional centrifugation (30 000 g, 4°C, 40 min) and passed first through a glass filter of pore size 10–16 μ m and then through a Membrex TF filter of pore size 0.2 μ m (MembraPure, Lörrweiler, Germany). The volume was reduced 10-fold by concentration with Amicon YM 10 membranes (Amicon, Witten, Germany). The concentrated supernatant was clarified by centrifugation and thoroughly dialyzed against 50 mM Tris-HCl, 1 M NaCl, pH 7.0 at 4°C. Immobilized metal affinity chromatography (IMAC) was performed at 4°C using a 5 ml column of Chelating Sepharose (Pharmacia) charged with Ni²⁺ and equilibrated with 50 mM Tris-HCl, 1 M NaCl, pH 7.0 (start buffer). The sample was loaded by passing the sample over the column. It was then washed with 20 column volumes of start buffer followed by start buffer containing 50 mM imidazole until the absorbance (280 nm) of the effluent was minimal (about 30 column volumes). Absorbed material was eluted with 50 mM Tris-HCl, 1 M NaCl, 250 mM imidazole, pH 7.0. After buffer exchange to 50 mM MES, pH 6.0, the protein was further purified on a Mono S ion-exchange column (Pharmacia). The purified scFv was dialyzed into PBS (15 mM sodium phosphate, 0.15 M NaCl, pH 7.4). For long-term storage, scFv were frozen in presence of BSA (final concentration 10 mg/ml) and kept at –80°C, as recommended (Kipriyanov *et al.*, 1995).

Isolation of scFv from inclusion bodies of bacteria transformed with plasmid pOPE51 was performed essentially as described previously (Kipriyanov *et al.*, 1996a).

SDS-PAGE and Western blot analysis

SDS-PAGE was carried out according to Laemmli (1970) under reducing conditions. Immunoblot analysis using anti *c-myc* mouse mAb 9E10 (Cambridge Research Biochemicals, Cambridge, UK) was performed as described previously (Kipriyanov *et al.*, 1994).

Analyses of scFv stability

For stability analyses, scFv preparations were stored at 4°C at a concentration 50 μ g/ml in PBS for 1 month. The activities of samples after storage were determined by flow cytometry.

Flow cytometry

We incubated 5 \times 10⁵ CD3⁺ Jurkat or CD3[–] JOK-1 cells in 50 μ l RPMI 1640 medium (Gibco BRL, Eggenstein, Germany) supplemented with 10% fetal calf serum (FCS) and 0.1% sodium azide (referred to as complete medium) with 100 μ l of a sample containing scFv for 45 min on ice. After washing with complete medium, the cells were incubated with 100 μ l of 10 μ g/ml anti *c-myc* mAb 9E10 (ICI Biochemicals) in the same buffer for 45 min on ice. After a second washing cycle, the cells were incubated with 100 μ l of FITC-labeled goat anti-mouse IgG (Gibco BRL) under the same conditions as before. The cells were then washed again and resuspended in 100 μ l of a 1 μ g/ml solution of propidium iodide (Sigma, Deisenhofen, Germany) in complete medium to exclude dead cells. The relative fluorescence of stained cells was measured using a FACScan flow cytometer (Becton Dickinson, Mountain View, CA).

Measurement of binding affinity

Affinities were derived either from the FACScan analysis of direct binding of scFv to Jurkat cells as described by Chamow *et al.* (1994) or from a competitive inhibition assay. In the

latter case, increasing concentrations of scFv were added to a subsaturating concentration of FITC-labeled mAb OKT3 (7.4 nM) and were incubated with Jurkat cells as described above. Fluorescence intensities of stained cells were measured as described above. Binding affinities were calculated according to the following equation derived from that of Schodin and Kranz (1993):

$$K_{a(I)} = (1 + [\text{FITC-OKT3}] \times K_{a(\text{OKT3})}) / \text{IC}_{50}$$

where I is the unlabeled inhibitor (scFv), [FITC-OKT3] is the concentration of FITC-labeled mAb OKT3, $K_{a(\text{OKT3})}$ is the binding affinity of mAb OKT3 ($1.2 \times 10^9 \text{ M}^{-1}$; Adair *et al.*, 1994) and IC_{50} is the concentration of inhibitor that yields 50% inhibition of binding.

Determination of the yield of soluble antibody fragments

The expression levels of soluble scFv fragments were determined in cleared culture medium and in crude periplasmic extracts isolated from shake-tube mini-cultures (5 ml). Culture supernatants were concentrated 20-fold using an Ultrafree-15 Biomax-10 centrifugal filter device (Millipore, Bedford, MA, USA) and dialyzed into PBS. The periplasmic extracts from cell pellets were prepared as described previously (Kipriyanov *et al.*, 1996b). For each scFv variant, three independent expression cultures were used. The concentrations of functional recombinant antibody fragments were determined from the fluorostaining of Jurkat cells using samples of periplasmic preparations and concentrated culture medium by the interpolation of their mean fluorescence intensities on the standard curves obtained with purified scFv of known concentration. At least four dilutions of samples were used for calculations.

Molecular modeling

Modeling was performed using AbM (Oxford Molecular, Oxford, UK). The framework was built by homology using HyHEL-5 (Sheriff *et al.*, 1987) for the parent light chain and 36-71 (Strong *et al.*, 1991) for the heavy chain. The complementarity determining regions (CDR) L1, L2, L3, H1 and H2 were built using canonical classes as proposed by Chothia *et al.* (1989) (CDR-L1 = Class 1, CDR-L2 = Class 1, CDR-L3 = Class 1, CDR-H1 = Class 1, CDR-H2 = Class 2) while CDR-H3 was built using the CAMAL algorithm (Martin *et al.*, 1989).

AbM sometimes has problems with junction regions where loops are spliced on to the framework. This can result in trigonal planar or D-amino acids at these junction sites. This occurred for residue H102 and this residue was rebuilt manually as an L-amino acid.

Other methods

Protein concentrations were determined by the Bradford dye-binding assay (Bradford, 1976) using the Bio-Rad protein assay kit (Bio-Rad Laboratories, Munich, Germany). The concentrations of purified scFv were calculated from the A_{280} values using the extinction coefficient $\epsilon^{1 \text{ mg/ml}} = 1.84$ derived from the Trp, Tyr and Phe content of the molecule using DNAid+1.8 Sequence Editor for Macintosh (F.Dardel and P.Bensoussan, Laboratoire de Biochimie, Ecole Polytechnique, Palaiseau, France). Analytical gel filtration of the scFv preparation was performed in PBS using a Superdex 75 HR10/30 column (Pharmacia). The sample volume and flow rate were 200 μl and 0.5 ml/min, respectively. For calibration of the column, a Low Molecular Weight Gel Filtration Calibration Kit (Pharmacia) was used.

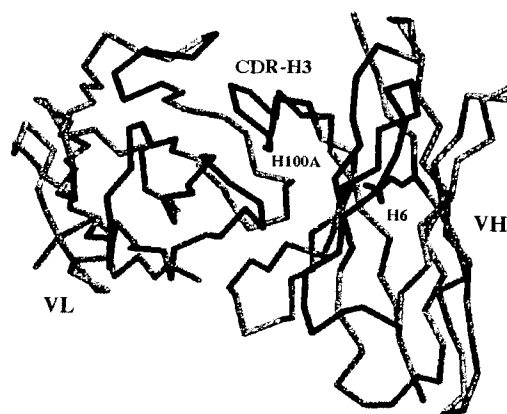


Fig. 1. Critical features of the OKT3 antigen binding site. The molecular model of OKT3 Fv is shown as a C α trace with side chains of amino acid residues H100A and H6.

Results

Modification of PCR amplified OKT3 V_H gene

The V region genes of the murine monoclonal antibody OKT3 (Van Wauwe *et al.*, 1980) were amplified by PCR from a cDNA preparation using two pairs of highly degenerate primers. Ten clones of each amplified V_H and V_L gene were sequenced and found to be identical. All the analyzed heavy chain variable regions contained QVQLQE as the N-terminal sequence. This region was encoded by the 5' primer Bi3f that contained a degenerate codon corresponding either to glutamic acid or glutamine residue at position 6 (Gotter *et al.*, 1995; Kipriyanov *et al.*, 1996b). The OKT3 scFv gene was assembled in the plasmid pOPE51 (Kipriyanov *et al.*, 1994) and expressed in *E. coli*. The resulting recombinant scFv product contained an unpaired cysteine residue near to its C-terminus that was specially introduced for making bivalent antibodies by chemical conjugation or for site-specific biotinylation (Kipriyanov *et al.*, 1994).

FACSscan analysis demonstrated no binding of scFv-OKT3 isolated from periplasmic inclusion bodies (Kipriyanov *et al.*, 1994; Kipriyanov *et al.*, 1995) to CD3-positive Jurkat cells (data not shown). A detailed analysis of the predicted structure based on the OKT3 V domain sequences allowed us to identify two amino acid residues in the V_H domain that might be critical for the activity of this recombinant antibody (Figure 1). First, a comparison with the OKT3 cDNA sequence (Adair *et al.*, 1994) showed that position 6 of FR1 was occupied by glutamine but not by glutamic acid as in the PCR amplification product. Furthermore, the consensus sequences of the Kabat database demonstrated that the cloned V_H gene fragment belongs to mouse immunoglobulin subgroup IIb (Kabat *et al.*, 1991), in which 92% of the members have Q in position 6. Second, the OKT3 V_H domain was found to contain a cysteine residue in the CDR-H3 which could interfere with folding by disrupting normal disulfide bonding, or might be oxidized during IMAC on an Ni column under denaturing conditions (Kipriyanov *et al.*, 1994). Therefore, we performed site-specific mutagenesis of the V_H gene to substitute E6 by Q and C100A [numbering scheme of Kabat *et al.* (1991)] by S. This double-mutant scFv-dmOKT3 (Q,S) demonstrated strong binding to CD3-positive Jurkat cells and no interaction with CD3-negative JOK-1 cells when purified from inclusion bodies (data not shown).

Construction and expression of anti-CD3 scFv mutants

To clarify how the amino acid changes described above contribute to the activity of the anti-CD3 scFv, we investigated four different scFv variants: a variant containing E6 and C100A that was amplified from hybridoma cDNA by PCR (E,C), a variant corresponding to the cDNA sequence published for OKT3 (Q,C; Adair *et al.*, 1994) and two variants containing Ser instead of Cys at V_H position 100A (E,S and Q,S).

To avoid working with inclusion bodies, which have to be refolded and to prevent vector-derived C-terminal unpaired cysteines from affecting the scFv properties (e.g. possible formation of an additional intramolecular disulfide bond with Cys-100A or scFv dimerization), we chose the plasmid pHOG21 for expressing the mutated scFv genes (Figure 2A). The bacterial pHOG21 expression vector was designed for the high-level production of soluble recombinant antibody fragments in *E. coli* (Kipriyanov *et al.*, 1996b). The antibody V_H fragment is preceded by a *pelB* leader sequence for secretion of recombinant antibody into the periplasmic space. The C-terminus of the V_H domain and N-terminus of the V_L domain are joined by a flexible 17 amino acid tag-linker that includes a tubulin epitope recognized by mAb YOL1/34 (Breitling *et al.*, 1991). A short peptide tag containing an epitope of the proto-oncogene *c-myc* recognized by mAb 9E10 (Evan *et al.*, 1985) is located at the C-terminus of the V_L domain followed by six histidine residues to facilitate the isolation of recombinant antibody fragments by IMAC. The sequence of the OKT3 derived scFv assembled in the plasmid pHOG21 is shown in Figure 2B with the mutations at amino acid positions 6 and 100A of the heavy chain indicated.

Recently, we showed that the addition of 0.4 M sucrose to the growth medium gives a 15–25-fold increase in the yield of soluble scFv for bacterial shake-tube cultures and an 80–150-fold increase for shake-flask cultures (Kipriyanov *et al.*, 1997). We also found that the scFv could be made to accumulate in the periplasm or be secreted into the medium by simply changing the incubation conditions and the concentration of the inducer. Therefore, to obtain higher yields of soluble anti-CD3 antibody fragments, we incubated induced *E. coli* cells in the presence of 0.4 M sucrose. Western blot analysis of cell pellets and periplasmic extracts of bacterial cultures expressing the four variants of OKT3 derived scFv demonstrated substantial differences in the ratio of soluble and total scFv (Figure 3). While the total amount of recombinant product found in the cell pellet seemed to be equal for all scFv variants, much less soluble scFv was found for variants containing Glu at position 6 (Figure 3, lanes 2 and 6). FACSscan analysis demonstrated the specific binding of periplasmic extracts for all the anti-CD3 scFv variants to CD3-positive Jurkat cells, although the fluorescence intensity obtained for scFvs with E6 was significantly lower (Figure 4A).

Purification of anti-CD3 scFv variants

To clarify whether the difference in antigen binding activity of periplasmic extracts containing different scFv variants (Figure 4A) is due to the difference in affinity or merely reflects the production levels of soluble antibody fragment, we performed a large-scale isolation of scFv using shake-flask bacterial cultures in the presence of 0.4 M sucrose. Under these conditions, we previously found that most of the secreted scFv was released into the medium (Kipriyanov *et al.*, 1997). The supernatant and periplasmic content of the induced bacterial culture was concentrated and passed through an Ni²⁺-

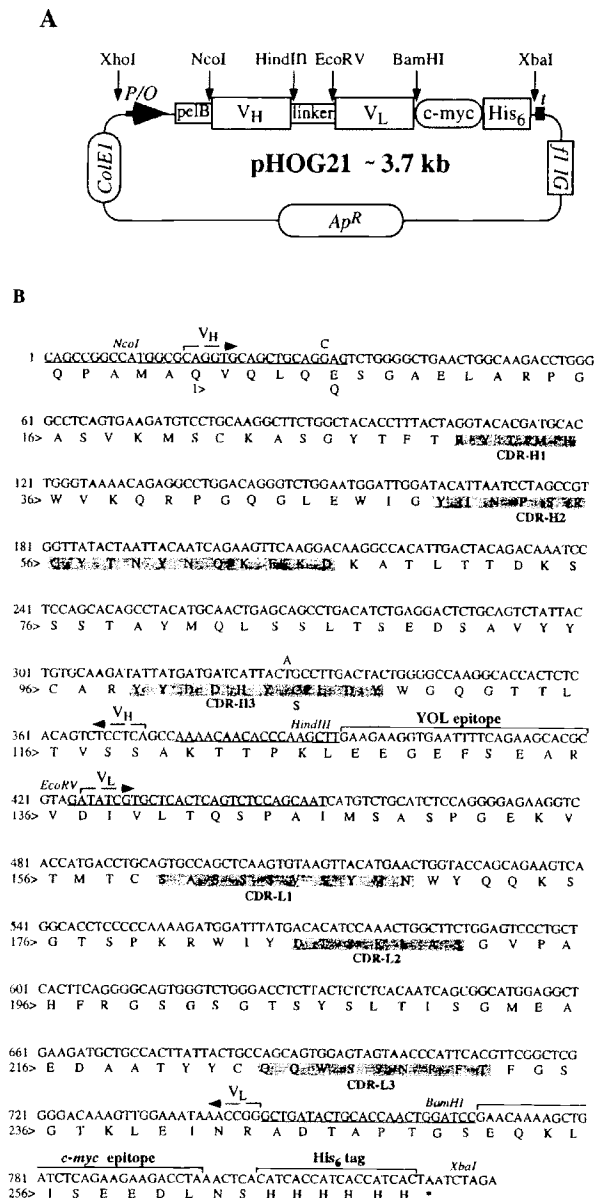


Fig. 2. The structure of scFv and expression vector. (A) Schematic representation of the plasmid pHOG21. Ap^R, ampicillin resistance-encoding gene; *c-myc*, a sequence encoding an epitope recognized by the monoclonal antibody 9E10; *ColE1*, origin of DNA replication; *f1*, intergenic region of phage f1; His₆, a sequence encoding six C-terminal histidine residues; linker, a sequence encoding 17 amino acids connecting the V_H and V_L domains; *pelB*, signal peptide sequence of bacterial pectate lyase; P/O, wt *lac* promoter/operator. (B) The nucleotide and deduced amino acid sequences of the scFv derived from hybridoma OKT3. The amino acid sequences corresponding to the complementarity determining regions (CDR) are shown shaded. The nucleotide sequences corresponding to PCR primers are underlined. The sequences coding for the YOL epitope in the linker region as well as *c-myc* epitope and six histidines in the carboxy terminal part of the scFv are indicated.

charged Chelating Sepharose column. After washing the column with buffer containing 50 mM imidazole, the bound scFv was eluted with 250 mM imidazole as a single peak in 2.5 column volumes. This purification procedure allowed us to isolate scFv in one step with a purity of about 95% (Figure

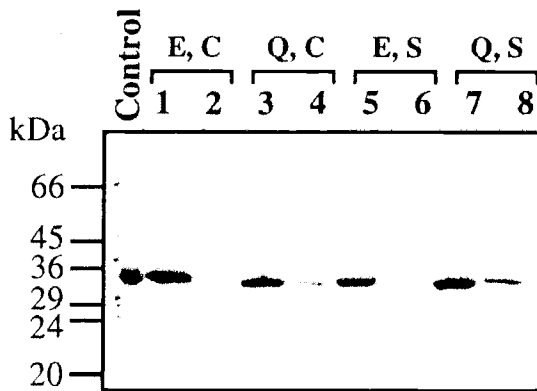


Fig. 3. Western blot analysis of cell pellets and periplasmic extracts from *E. coli* clones expressing different anti-CD3 scFv variants. Lanes: 1, 3, 5, 7, total cell lysate from induced bacteria corresponding to 100 μ l of culture; 2, 4, 6, 8, periplasmic extracts corresponding to 180 μ l of culture. The scFv were detected using mAb 9E10 recognizing the C-terminal *c-myc* epitope. As a control, 1 μ g of pure scFv-dmOKT3 isolated from inclusion bodies was used. The positions of molecular mass markers are shown on the left.

5A). The main contaminant present in samples of scFv purified by IMAC has recently been identified as an *E. coli* metal-binding 27 kDa WHP protein (Wülfing *et al.*, 1994). An analysis of its amino acid composition showed that the WHP protein has an isoelectric point (*Ia*) of 5.16; anti-CD3 scFv variants were found to be more basic (the calculated *Ia* was between 7.27 for the E,C and 7.52 for the Q,S variant). This charge difference allowed us to purify the recombinant antibody fragments to homogeneity by ion-exchange chromatography on a Mono S column (Figure 5B). Analytical gel-filtration on a Superdex 75 column demonstrated that all the isolated scFv preparations consisted only of monomers (data not shown).

Affinity and stability measurements

Our attempts to use radioiodinated scFv preparations for measuring the direct binding of recombinant antibodies to CD3-positive Jurkat cells were unsuccessful. Unfortunately, iodination using chloramine-T yielded an inactive product for both anti-CD3 scFv and Fab fragment prepared from mAb OKT3 (data not shown). It is possible that iodination blocked tyrosine residues in the CDR regions that may be important for antigen-binding (Figure 2B). We therefore employed two different non-radioactive approaches based on flow cytometry (Bohn, 1980) that do not require any modification of the protein.

In the first approach, recombinant antibody fragments were incubated with cells as in a standard radioprotein binding assay, except that an anti-*c-myc* mAb and fluorescent anti-mouse IgG reagent were used to detect the amount of bound scFv. In comparison with a standard radioligand binding assay, the same variables (except the number of molecules bound at saturation) can be measured and an affinity constant determined from the slope of the resultant Scatchard curve (Chamow *et al.*, 1994).

The binding of scFv preparations was measured using human Jurkat cells as a source of naturally expressed cell bound CD3 ϵ . Binding to CD3-negative JOK-1 cells was used as a negative control. The results of fluorostaining of Jurkat cells displayed in Figure 4B demonstrate that the same concentrations of different scFv variants yield similar fluorescence (slightly higher values were obtained for variants containing Gln at position 6). A pattern of increased fluorescence with

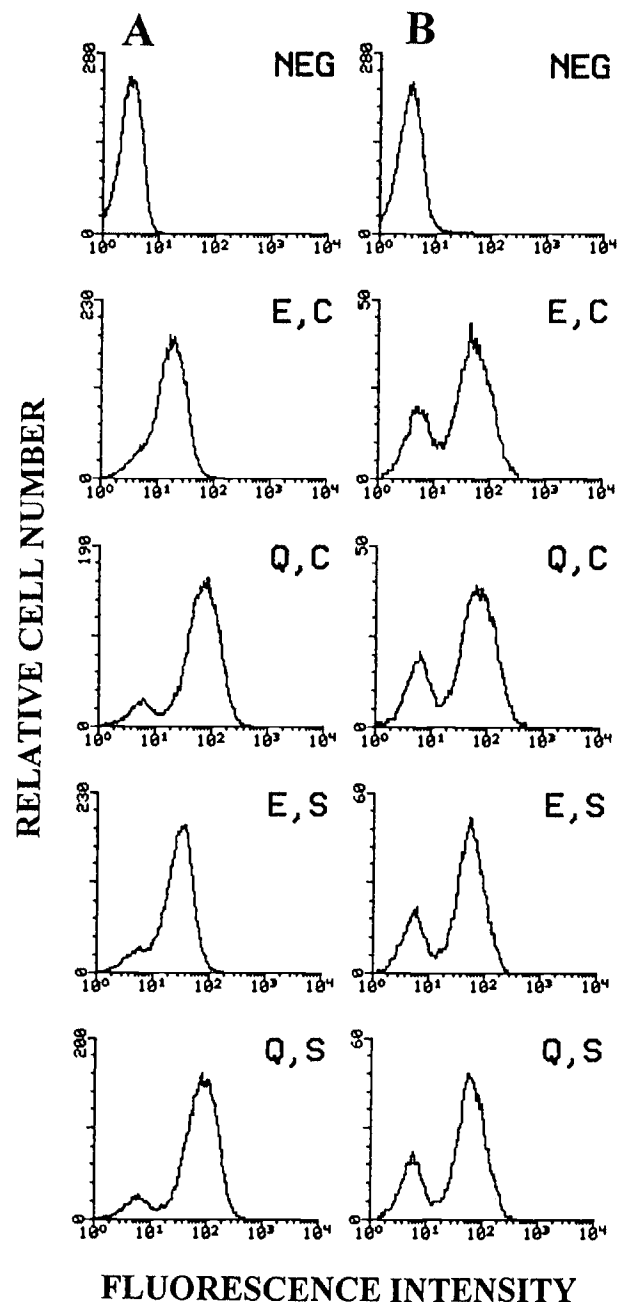


Fig. 4. Flow cytometric analysis of the binding of anti-CD3 scFvs to Jurkat cells. (A) Analysis of binding of periplasmic extracts. (B) Analysis of binding of pure scFv preparations at concentration 25 μ g/ml. The presence of two peaks of fluorescence indicates that not all cells of the used line express CD3 antigen. As a negative control, binding to CD3-negative JOK-1 was used.

increased amounts of scFv was observed that seems to reach a plateau at higher concentrations (Figure 6A). On the basis of fluorescence measurements at different concentrations of added scFv, typical Scatchard curves were generated from which K_a values were derived (data not shown).

In a second approach, the binding efficiency of anti-CD3 scFv variants to Jurkat T cells was investigated by competition

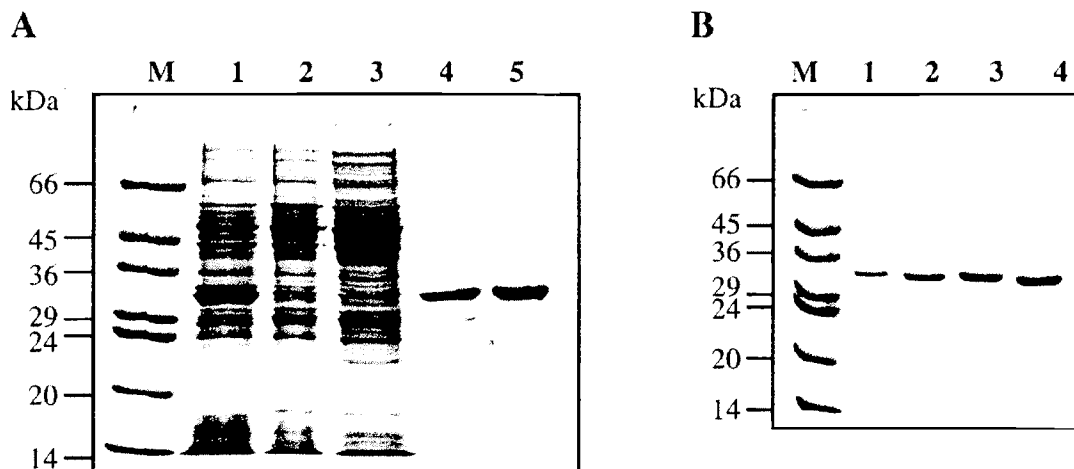


Fig. 5. 12% SDS-PAGE analysis of scFv preparations. (A) Analysis of scFv-dmOKT3 (Q,S) at different steps of purification. Lanes: M, molecular mass markers (values in kDa are shown on the left); 1, total cell lysate; 2, soluble periplasmic content; 3, concentrated culture medium; 4, scFv isolated by IMAC; 5, scFv purified by ion-exchange chromatography. (B) Analysis of purified scFv preparations for variant E,C (lane 1), Q,C (lane 2), E,S (lane 3) and Q,S (lane 4). The gels were stained with Coomassie Brilliant Blue.

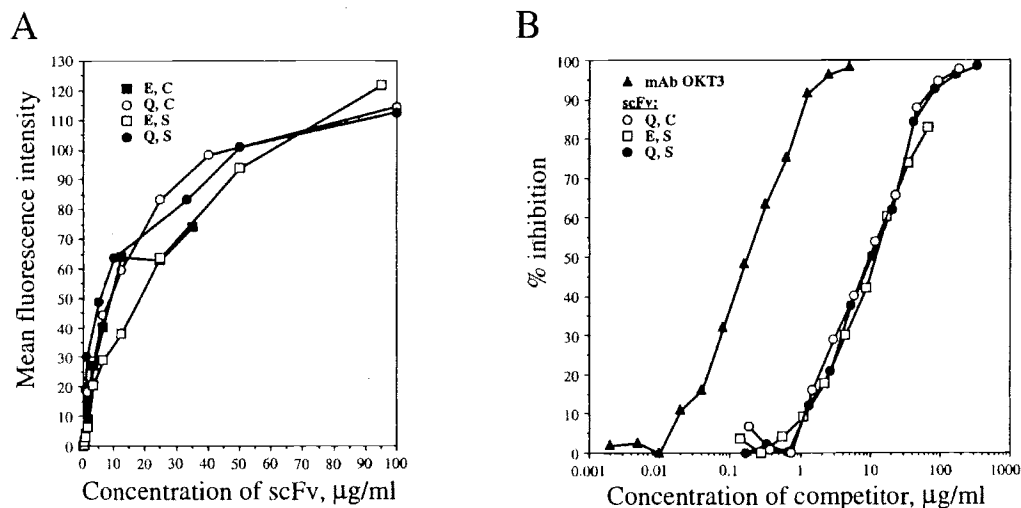


Fig. 6. Analyses of apparent affinities of scFv variants by flow cytometry. (A) Direct binding of different scFv variants to Jurkat cells. The cells were incubated with various concentrations of OKT3 derived scFv variants: E,C (filled squares), Q,C (open circles), E,S (open squares) and Q,S (filled circles). The curves were constructed by plotting the mean fluorescence intensity against the scFv concentration. (B) Inhibition of binding of FITC-OKT3 to Jurkat cells in presence of mAb OKT3 (filled triangles) and scFv variants: Q,C (open circles), E,S (open squares) and Q,S (filled circles).

with FITC-labeled mAb OKT3. The data presented in Figure 6B demonstrate that the OKT3 derived scFv Q,C, E,S and Q,S variants competed similarly, all at ~100 times the concentration of the intact IgG OKT3.

Analysis of the stability of anti-CD3 scFv variants after storage in PBS for 1 month at 4°C demonstrated a substantial loss of antigen-binding activity for scFv containing Cys in CDR-H3 (Figure 7).

Table I summarizes the results of the affinity and stability measurements. The apparent affinity values obtained for all the scFv variants proved to be quite close, indicating (i) only a slight effect of the sixth amino acid on the antigen binding and (ii) that the replacement of Ser for Cys in the middle of CDR-H3 does not disturb the antigen-antibody complex. Both the glutamic acid at position H6 and especially the cysteine at position of H100A led to a decreased stability of the scFv,

probably because of a higher tendency for such antibody fragments to aggregate and/or for oxidation of unpaired cysteine residues during storage. No proteolytic degradation during storage was detected for any of the examined scFv variants (data not shown).

Analysis of expression yields

To study the influence of positions H6 and H100A on the production levels of soluble scFv fragments, we analyzed the antigen binding activities of periplasmic extracts and the concentrated culture medium of bacteria expressing scFv E,C, Q,C, E,S and Q,S variants. The expression yield data presented in Table I demonstrate that a single amino acid substitution of E6 by Q yields more than a 30-fold increase in soluble scFv product. In contrast, the single exchange of C100A by S led to a more moderate twofold increase in soluble scFv. These

effects were cumulative: the total yield of the Q,S variant was 66-fold higher than that for the E,C scFv variant (Table I). For all the examined variants, a small proportion of the functional soluble scFv was found to be released into the culture medium.

Discussion

Recombinant antibody fragments directed against cell surface antigens can provide useful components for the development of therapeutic agents. To target cytotoxic effector T cells to a tumor site, we have constructed an anti-human CD3 single-chain antibody by PCR amplification of the immunoglobulin variable domain genes from cDNA of the hybridoma OKT3. Expression of the assembled scFv gene in *E.coli* yielded a non-functional product after refolding from inclusion bodies.

In general, the primers we and other workers use for amplifying V genes from hybridoma cDNAs are designed to match all the known sequences of immunoglobulin genes. However, PCR amplification using degenerate primers does

not always yield a gene with naturally occurring codons in the primer region (McCartney *et al.*, 1995). It is therefore often not possible to know which codons occur naturally if, as in our case, the DNA sequence was not then available. For example, the same set of primers resulted either in Glu or Gln in H6 after amplification of the V_H gene of an antibody against anti-human CD19 (Kipriyanov *et al.*, 1996b). Regarding the significance of this position, there was no indication in the literature that it may be critical for bacterially expressed antibody fragments.

To improve the properties of the recombinant antibody fragment, we focused on the amino acid residues which are structurally uncommon for the V_H subgroup IIb: glutamic acid at the position H6 of FR1 and a cysteine in the middle of the CDR-H3 loop. Site-specific mutagenesis and a change of expression system (soluble secreted scFv versus inclusion bodies) allowed us to clarify their influence on the production of a functional scFv antibody fragment.

We demonstrated that a single amino acid substitution of E by Q at position 6 of the heavy chain resulted in a 30-fold increase in soluble scFv product and significantly increased the stability of the recombinant molecule during storage. However, this substitution had very little effect on the affinity (scFv containing Q had affinity constants about 1.5 times higher than variants with E). This slight difference may be explained by the possible difference in the percentage of functional scFv (Kipriyanov *et al.*, 1994). We can therefore conclude that the sixth amino acid influences the correct folding of the V_H domain, perhaps by affecting some folding intermediate, but it has little or no effect on antigen binding. This conclusion was supported by computational molecular modeling. Examination of the residues which surround position 6 of the heavy chain in the three-dimensional model reveals no reason why a Glu or Gln residue should have any significant effect on the conformations of the CDRs (Figure 1).

It is not clear how Glu may effect the folding of the scFv fragment in the bacterial environment because very little is known about this process. Attempts have been made to prevent the side reaction of aggregation by overexpressing some known enzymes of the *E.coli* folding machinery such as the GroES/L chaperones, disulfide-isomerase and proline-*cis-trans*-isomerase (Knappik *et al.*, 1993; Duenas *et al.*, 1994). However, these proteins did not increase the yield of soluble antibody fragments. The presence of a periplasmic chaperone has therefore been postulated but not yet identified (Wülfing and Plückthun, 1994). From a variety of experiments, evidence is accumulating that the primary sequence of the antibody

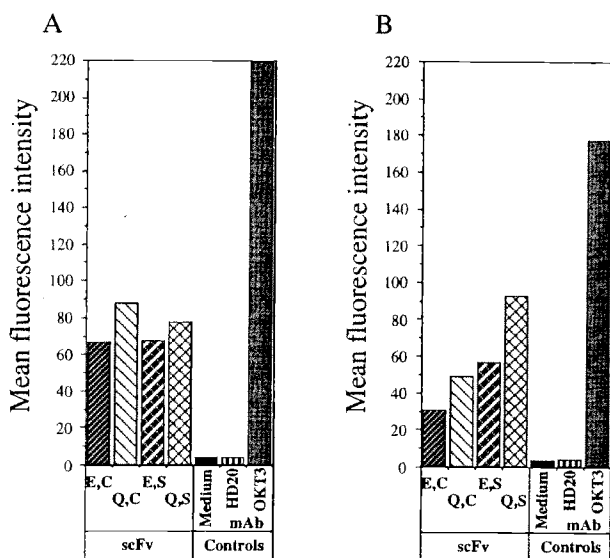


Fig. 7. Flow cytometric analyses of interaction of OKT3 derived scFv variants with Jurkat cells. (A) Fluorescence intensity obtained for fresh scFv preparations at concentration 25 µg/ml. (B) Fluorescence intensity obtained for the same scFv preparations after storage in PBS for 1 month at 4°C. As controls, the interaction of culture medium, mAb OKT3 and irrelevant mAb HD20 with Jurkat cells is shown.

Table I. Expression levels of anti-CD3 scFv variants, their stabilities and affinities to human CD3 antigen

| scFv variant | Yield of scFv (µg/l of culture) ^a | Release of scFv (%) ^b | K _d ^c (M ⁻¹ /10 ⁷) | K _d ^d (M ⁻¹ /10 ⁷) | Stability (%) ^e |
|--------------|--|----------------------------------|---|---|----------------------------|
| E,C | 72.7 ± 19.5 ^f | 9.6 ± 1.4 ^f | 1.09 | n.d. ^g | 37.87 |
| Q,C | 2314.7 ± 578.8 | 24.4 ± 1.7 | 1.96 | 3.16 | 46.63 |
| E,S | 148.4 ± 37.7 | 25.4 ± 7.9 | 1.27 | 2.49 | 70.10 |
| Q,S | 4846.0 ± 477.3 | 29.4 ± 1.9 | 1.42 | 2.95 | 100 |

^aTotal amount of soluble scFv both in crude periplasmic extract and culture medium estimated by flow cytometry.

^bPercentage of total scFv amount found in culture medium.

^cBinding constants as determined by cytofluorometric Scatchard analysis.

^dBinding constants as determined from a competitive inhibition assay using FITC-OKT3.

^eActivity (%) after 1 month at 4°C as determined by flow cytometry.

^fArithmetic mean and standard deviation based on three independent experiments.

^gNot determined.

plays a decisive role in the efficiency of folding in a bacterial environment (Carter *et al.*, 1992; Knappik and Plückthun, 1995). Our own results lead to a similar conclusion. The amino acid H6 influences not only the folding efficiency but also the stability of the correctly folded scFv.

The single exchange of Cys at position H100A by Ser also led to a twofold increase in soluble scFv. Three residues before the start of CDR-H3 is a conserved cysteine at position H92 which forms a structural disulfide bond with position H22. Thus, having another Cys nearby (at H100A) could easily allow mis-folding where H100A instead of H92 is involved in forming the disulfide bond with H22, thereby generating a mis-folded, insoluble and non-functional product. Analogously, Ostermeier *et al.* (1995) demonstrated that substitution of an uncommon cysteine at position H50 (first amino acid of CDR-H2) by a serine led to a 20-fold increase in soluble Fv production. Since the authors of this work were working with an Fv fragment, a mutation in the V_H domain can only influence the yield of the heavy chain fragment. Although direct comparisons of the influence of Cys residues in different CDRs on correct folding cannot be made, these results suggest that such uncommon residues may be more critical for the folding of a single antibody domain (V_H) than for scFv.

In our case, a cysteine was substituted that is present directly in the middle of CDR-H3, which is in the middle of the antigen-combining site and generally has the greatest influence on binding affinity. CDR-H3 plays a prominent role not only in ligand binding, but also in the contact with the V_L domain and with the other CDRs (Padlan, 1994). Although cysteine can occasionally form hydrogen bonds, this is rare in proteins (Baker and Hubbard, 1984) and it is a relatively hydrophobic residue. We therefore considered two possible mutations at H100A: serine (maintaining the size as closely as possible, but introducing a very hydrophilic residue) and valine (increasing the hydrophobic nature, but adding an extra atom). Given that the residue is exposed to solvent in the model, we chose to make the mutation to serine since any increase in hydrophobicity could lead to a change in folding of the loop. We were aware, however, that the substitution might interfere with antigen binding or influence the contact between the variable domains. Fortunately, the Cys to Ser mutation had no effect on antigen binding and, as hoped, led to a significant improvement in the stability of the scFv. Although the exposure patterns of the various amino acid types in immunoglobulins are comparable to those in other water-soluble proteins, cysteines are more exposed in CDRs than they are in the framework regions (Padlan, 1994). This is especially true for short (10 residues or less) hypervariable loops which do not have much opportunity to bury one of their residues while maintaining a distorted hairpin conformation for antigen binding. Exposure of the cysteine 100A SH group to solvent may result in oxidation or modification of the group over time and this may have an influence on the stability of the antigen-antibody complex. It is also possible that the unpaired cysteines of two adjacent scFv molecules could form a disulfide bond, thus giving rise to inactive and probably insoluble scFv dimers and causing a decrease in the concentration of functional scFv. These factors would explain the experimentally observed instability during the storage of scFv variants containing Cys in CDR-H3.

It is worth noting that in the present work we actually compared two different strategies of folding, i.e. *in vivo* and *in vitro*. The renaturation procedure, which has been used to

refold several antibody fragments (Kipriyanov *et al.*, 1994; Gotter *et al.*, 1995) and a more complex scFv:streptavidin fusion protein (Kipriyanov *et al.*, 1996a) did not lead to the formation of an active scFv-E₂C variant. These results point to limitations in the folding strategy *in vitro* compared with *in vivo* and indicate how such problems can be overcome.

In conclusion, we have constructed a modified version of an anti-human CD3 scFv antibody fragment with improved stability *in vitro* and increased production level in bacteria. This molecule may be particularly useful for the creation of recombinant cytotoxic immunoconjugates.

Acknowledgements

S.M.K. was supported by a grant from the Deutsche Krebshilfe/ Mildred Scheel Stiftung and by the Fond der Deutschen Chemischen Industrie. A.C.R.M. was supported by the UK Medical Research Council and O.A.K. by a fellowship from Volkswagen Foundation.

References

- Adair, J.R. *et al.* (1994) *Hum. Antibod. Hybridomas*, **5**, 41–47.
- Adams, G.P. *et al.* (1993) *Cancer Res.*, **53**, 4026–4034.
- Ayala, M., Balint, R.F., Fernandez-de-Cossio, M.E., Canaan-Haden, L., Larrick, J.W. and Gavilondo, J.V. (1995) *Biotechniques*, **18**, 832–842.
- Baker, E.N. and Hubbard, R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
- Better, M., Chang, C.P., Robinson, R.R. and Horwith, A.H. (1988) *Science*, **240**, 1041–1043.
- Bird, R.E. *et al.* (1988) *Science*, **242**, 423–426.
- Bohlen, H., Hopff, T., Mancke, O., Engert, A., Kube, D., Wickramanayake, P.D., Diehl, V. and Tesch, H. (1993) *Blood*, **82**, 1803–1812.
- Bohn, B. (1980) *Mol. Cell. Endocrinol.*, **20**, 1–15.
- Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
- Breitling, F., Dübel, S., Seehaus, T., Klewinghaus, I. and Little, M. (1991) *Gene*, **104**, 147–153.
- Carter, P. *et al.* (1992) *BioTechnology*, **10**, 163–167.
- Chamow, S.M., Zhang, D.Z., Tan, X.Y., Mhatre, S.M., Marsters, S.A., Peers, D.H., Bym, R.A., Ashkenazi, A. and Junghans, R.P. (1994) *J. Immunol.*, **153**, 4268–4280.
- Chothia, C. *et al.* (1989) *Nature*, **342**, 877–883.
- Dübel, S., Breitling, F., Fuchs, P., Zewe, M., Gotter, S., Welsch, M., Moldenhauer, G. and Little, M. (1994) *J. Immunol. Methods*, **175**, 89–95.
- Duenas, M., Vazquez, J., Ayala, M., Söderlind, E., Ohlin, M., Perez, L., Borrebaeck, C.A.K. and Gavilondo, J.V. (1994) *BioTechniques*, **16**, 476–483.
- Duenas, M., Ayala, M., Vazquez, J., Ohlin, M., Söderlind, E., Borrebaeck, C.A.K. and Gavilondo, J.V. (1995) *Gene*, **158**, 61–66.
- Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, M. (1985) *Mol. Cell. Biol.*, **5**, 3610–3616.
- Gotter, S., Haas, C., Kipriyanov, S., Dübel, S., Breitling, F., Khazaie, K., Schirmacher, V. and Little, M. (1995) *Tumor Targeting*, **1**, 107–114.
- Huston, J.S. *et al.* (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5879–5883.
- Kabat, E.A., Wu, T.T., Perry, H.M., Gottesmann, K.S. and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, 5th edn. Public Health Service, National Institutes of Health, Bethesda, MD.
- Kipriyanov, S.M., Dübel, S., Breitling, F., Kontermann, R.E. and Little, M. (1994) *Mol. Immunol.*, **31**, 1047–1058.
- Kipriyanov, S.M., Dübel, S., Breitling, F., Kontermann, R.E., Heymann, S. and Little, M. (1995) *Cell Biophys.*, **26**, 187–204.
- Kipriyanov, S.M., Little, M., Kropshofer, H., Breitling, F., Gotter, S. and Dübel, S. (1996a) *Protein Engng.*, **9**, 203–211.
- Kipriyanov, S.M., Kupriyanova, O.A., Little, M. and Moldenhauer, G. (1997) *J. Immunol. Methods*, **196**, 51–62.
- Kipriyanov, S.M., Moldenhauer, G. and Little, M. (1997) *J. Immunol. Methods*, **200**, 69–77.
- Knappik, A., Krebber, C. and Plückthun, A. (1993) *BioTechnology*, **11**, 77–83.
- Knappik, A. and Plückthun, A. (1995) *Protein Engng.*, **8**, 81–89.
- Kung, P.C., Golstein, G., Reinherz, E.L. and Schlossman, S.F. (1979) *Science*, **206**, 347–349.
- Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
- Laemmli, U.K. (1970) *Nature*, **227**, 680–685.
- Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9268–9272.
- McCartney, J.E. *et al.* (1995) *Protein Engng.*, **8**, 301–314.

- Milenic,D.E., Yokota,T., Filpula,D.R., Finkelman,M.A.J., Dodd,S.W., Wood,J.F., Whitlow,M., Snoy,P. and Schlom,J. (1991) *Cancer Res.* **51**, 6363–6371.
- Nitta,T., Sato,K., Yagita,H., Okumura,K. and Ishii,S. (1990) *Lancet*, **335**, 368–376.
- Ostermeier,C., Essen,L.-O. and Michel,H. (1995) *Proteins: Struct. Funct. Genet.*, **21**, 74–77.
- Padlan,E.A. (1994) *Mol. Immunol.*, **31**, 169–217.
- Sanger,F., Nicklen,S. and Coulson,R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Sanna,P.P., De Logu,A., Williamson,R.A., Samson,M.E., Altieri,D.C., Bloom,F.E. and Burton,D.R. (1995) *BioTechnology*, **13**, 1221–1224.
- Schodin,B.A. and Kranz,D.M. (1993) *J. Biol. Chem.*, **268**, 25722–25727.
- Sheriff,S., Silverton,W.W., Padlan,E.A., Cohen,G.H., Smith-Gill,S.J., Finzel,B.C. and Davies,D.R. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8075–8079.
- Skerra,A. and Plückthun,A. (1988) *Science*, **240**, 1038–1041.
- Strong,R.K., Campbell,R., Rose,D.R., Petsko,G.A., Sharon,J., Margolies,M.N. (1991) *Biochemistry*, **30**, 3739–3748.
- Thistlethwaite,J.R., Cosimi,A.B., Delmonico,F.L., Rubin,R.H., Talkoff-Rubin,N., Nelson,P.W., Fang,L. and Russel,P.S. (1984) *Transplantation*, **38**, 695–701.
- Transy,C., Moingeon,P.E., Marshall,B., Stebbins,C. and Reinherz,E.L. (1989) *Eur. J. Immunol.*, **19**, 947–950.
- Van Wauwe,J.P., DeMey,J.R. and Goossens,J.G. (1980) *J. Immunol.*, **124**, 2708–2713.
- Winter,G. and Milstein,C. (1991) *Nature*, **349**, 293–299.
- Woodle,E.S., Thistlethwaite,J.R., Emond,J.C., Whittington,P.F., Black,D.D., Aran,P.P., Baker,A.L., Stuart,F.P. and Broelsch,C.E. (1991) *Transplantation*, **51**, 1207–1212.
- Woodle,E.S. *et al.* (1992) *J. Immunol.*, **148**, 2756–2763.
- Wülfing,C. and Plückthun,A. (1994) *J. Mol. Biol.*, **242**, 655–669.
- Wülfing,C., Lombardero,J. and Plückthun,A. (1994) *J. Biol. Chem.*, **269**, 2895–2901.
- Yannelli,J.R., Crumpacker,D.B., Good,R.W., Friddell,C.D., Poston,R., Horton,S., Maleckar,J.R. and Oldham,R.K. (1990) *J. Immunol. Methods*, **1**, 91–100.
- Yokota,T., Milenic,D.E., Whitlow,M. and Schlom,J. (1992) *Cancer Res.*, **52**, 3402–3408.

Received July 8, 1996; revised December 17, 1996; accepted December 30, 1996

Structural Effects of Framework Mutations on a Humanized Anti-Lysozyme Antibody¹

Margaret A. Holmes,* Timothy N. Buss,* and Jefferson Foote^{2*†}

A humanized version of the mouse anti-lysozyme Ab D1.3 was previously constructed as an Fv fragment and its structure was crystallographically determined in the free form and in complex with lysozyme. Here we report five new crystal structures of single-amino acid substitution mutants of the humanized Fv fragment, four of which were determined as Fv-lysozyme complexes. The crystals were isomorphous with the parent forms, and were refined to free *R* values of 28–31% at resolutions of 2.7–2.9 Å. Residue 27 in other Abs has been implicated in stabilizing the conformation of the first complementarity-determining region (CDR) of the H chain, residues 31–35. We find that a Phe-to-Ser mutation at 27 alters the conformation of immediately adjacent residues, but this change is only weakly transmitted to Ag binding residues in the nearby CDR. Residue 71 of the H chain has been proposed to control the relative disposition of H chain CDRs 1 and 2, based on the bulk of its side chain. However, in structures we determined with Val, Ala, or Arg substituted in place of Lys at position 71, no significant change in the conformation of CDRs 1 and 2 was observed. *The Journal of Immunology*, 2001, 167: 296–301.

Humanized Abs are created by replacing the complementarity-determining regions (CDRs)³ of a human Ab (as defined by Wu and Kabat; Refs. 1, 2) with the corresponding CDRs of a nonhuman Ab (3). This CDR graft transfers the antigenic specificity of the CDR donor molecule, but leaves the new engineered molecule immunologically human, inasmuch as the immunogenicity of humanized Abs in humans is extremely low (4, 5). The first humanized Ab was specific for the hapten nitrophenacetyl. This molecule had been CDR grafted in the H chain only, which was coexpressed with a mouse L chain. The humanized anti-nitrophenacetyl showed 1.5- to 3-fold reduced hapten affinity relative to a control molecule with murine sequences in both chains (3). This finding of altered affinity proved that framework residues can influence the structure of the Ag combining site. Riechmann et al. (4) confirmed this finding in a humanized anti-CD52. The initial humanized construct showed weak avidity. A single Ser-to-Phe mutation at framework residue H27⁴ restored avidity to near that of the fully murine control. The importance of framework residues in maintaining the structure of the CDRs and the frequent need for mutational revisions in the framework have since been confirmed many more times during the engineering of humanized Abs to have avidity matching that of their murine antecedents (6).

We developed a humanized anti-lysozyme (HuLys) as a model system for studying structural issues attending the transfer of CDRs from a murine to a human framework (7–9). Thus, murine and human segments for the construction were chosen from among Ab V domains whose structures had been determined. The six CDRs of HuLys come from the murine Ab D1.3, which was raised against hen egg lysozyme (10, 11). The structure of the D1.3 heterodimer of H and L chain V regions (Fv) has been determined at 1.8-Å resolution in both the liganded and unliganded forms (12, 13). The HuLys H chain framework (residues H1–H30, H36–H49, H66–H94, and H103–H113 in the Kabat numbering system) comes from the human myeloma protein NEW, whose structure has been determined at 2.0 Å (14). The κ L chain framework (residues L1–L23, L35–L49, L57–L88, and L98–L108) is a consensus sequence similar to that of the human Bence-Jones protein REI, also determined at 2.0 Å (15).

The crystal structures of the HuLys Fv in free form (16) and complexed with the Ag lysozyme (17) were previously determined at 2.9 and 2.7 Å, respectively (Brookhaven Protein Data Bank accession numbers 1BVL and 1BVK). In this work, we describe crystal structures of a series of single substitution mutants of the HuLys Fv, viz H27S, H71V, H71A, and H71R.

*Program in Molecular Medicine, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and [†]Department of Immunology, University of Washington, Seattle, WA 98195

Received for publication September 28, 2000. Accepted for publication April 26, 2001.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ This work was supported by the Department of the Army Breast Cancer Research Program (Grant DAMD 17-97-1-7124).

² Address correspondence and reprint requests to Dr. Jefferson Foote, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, C3-168, P.O. Box 19024, Seattle, WA 98109-1024. E-mail address: jfoote@fhcrc.org

³ Abbreviations used in this paper: CDR, complementarity-determining region; Fv, heterodimer of H and L chain V regions; HuLys, humanized anti-lysozyme; rms, root mean square.

⁴ Residues are numbered using the Kabat system and preceded by a chain designator, e.g., H71 for residue 71 in the H chain. The wild-type Fv has Phe at residue H27 and Lys at residue H71; mutant molecules are designated by the substitution, e.g., H71V is an Fv with Val at residue H71.

Materials and Methods

Protein engineering

Fvs were expressed in *Escherichia coli* using the pAK19 vector (18), which uses a *phoA* promoter and heat-stable enterotoxin II leader sequence. This vector directs gene products to the periplasm, from which correctly folded, disulfide-oxidized molecules are released after cell harvest. Material used in the present work was released from the periplasm by osmotic shock and purified by affinity chromatography on lysozyme-Sepharose, as described previously (17). Protein concentrations were determined spectrophotometrically, using calculated extinction coefficients (19).

Crystal growth

Crystals of the four mutant complexes were grown in the same way as the native complex crystals (17). Each of the HuLys Fv solutions was mixed with a lysozyme solution in equimolar proportions. The mixtures then sat from several hours to 2 days. PBS was added to dilute the solution, which was centrifuged before use. Protein concentrations ranged from 6.5 to 10.5 mg/ml. The reservoir for vapor diffusion was 0.8 M K₂HPO₄, 0.8 M

Table I. Data collection

| Structure | H27S Complex | | H71V Complex | | H71A Complex | | H71R Complex | | H71V | |
|--|--|-----------|--|-----------|--|-----------|--|-----------|--|-----------|
| Space group | P4 ₁ 2 ₁ 2 | | P4 ₁ 2 ₁ 2 | | P4 ₁ 2 ₁ 2 | | P4 ₁ 2 ₁ 2 | | P4 ₃ 2 ₁ 2 | |
| Fv/asymmetric unit | 2 | | 2 | | 2 | | 2 | | 2 | |
| Cell dimensions (Å) | <i>a</i> = <i>b</i> =97.9; <i>c</i> =173.3 | | <i>a</i> = <i>b</i> =96.3; <i>c</i> =175.7 | | <i>a</i> = <i>b</i> =97.6; <i>c</i> =174.1 | | <i>a</i> = <i>b</i> =97.1; <i>c</i> =174.8 | | <i>a</i> = <i>b</i> =146.8; <i>c</i> =71.9 | |
| Resolution (Å) | 50.0–2.7 | 2.75–2.70 | 50.0–2.7 | 2.75–2.70 | 50.0–2.7 | 2.75–2.70 | 50.0–2.7 | 2.75–2.70 | 50.0–2.9 | 2.95–2.90 |
| Measured reflections | 132,657 | >2580 | 141,035 | >3001 | 108,158 | >2271 | 141,201 | >2672 | 88,937 | >1849 |
| Unique reflections | 23,047 | 1060 | 22,318 | 1133 | 21,229 | 1044 | 22,707 | 1025 | 16,618 | 751 |
| Completeness (%) | 96.3 | 90.3 | 95.1 | 97.5 | 88.9 | 90.5 | 95.6 | 89.0 | 92.3 | 87.0 |
| <i>R</i> value ^a | 0.065 | 0.315 | 0.067 | 0.404 | 0.059 | 0.310 | 0.065 | 0.376 | 0.075 | 0.403 |
| Average <i>I</i> / <i>σ</i> ₁ | 10.1 | 2.6 | 18.1 | 2.3 | 17.4 | 2.6 | 16.5 | 2.3 | 17.2 | 2.5 |

$$^a R = \sum |I_{hkl} - \langle I_{hkl} \rangle| / \sum I_{hkl}$$

NaH₂PO₄, 0.1 M HEPES, pH 6.5. Sitting drops consisting of equal volumes of complex solution and reservoir solution were set up in microbridges.

Crystals of the uncomplexed H71V Fv were grown by macroseeding. The seeds were obtained from a hanging drop vapor diffusion crystallization that used 16 mg/ml protein and a reservoir of 0.74 M sodium citrate, 0.01% NaN₃, pH 6.5. Two rounds of seeding were performed. Each time, a few crystals were removed from the drop and placed in a microbridge in a fresh drop composed of equal volumes of Fv solution (16 mg/ml) and reservoir solution (0.8 M sodium citrate, 0.01% NaN₃, pH 6.5).

Data collection

X-ray diffraction data sets were collected from single crystals at 4°C using an *R* axis detector. The data sets were processed with DENZO and SCALEPACK (20, 21). Details of the processing are given in Table I. Before refinement, the data sets were partitioned into a working set and a test set. The test sets for the complexes contained only reflections that had made up the test set for the refinement of the native complex structure, so as to maintain the independence of the test set (22). The test set for the uncomplexed Fv was created by X-PLOR (23), as the refinement of the native Fv structure did not involve a test set.

Refinement

Refinement of the structure of the HuLys H27S Fv-lysozyme complex began with the model of the native complex with residue H27 changed to Gly. A round of rigid body refinement at 3.5-Å resolution was followed by rounds of positional refinement at 2.7-Å resolution using X-PLOR and model building of the loop containing the mutation. A cycle of torsion angle molecular dynamics refinement was run, followed by more rounds of positional refinement and model building. Omit map density was sufficient to model only one of the two H27 side chains. The refinement was completed with a cycle of individual *B* value refinement with TNT (24) and a cycle of X-PLOR *B* value refinement. Refinement statistics are given in Table II.

Refinement of the structures of the HuLys H71V, H71A, and H71R Fv-lysozyme complexes was more straightforward. The starting model was

the native complex with H71 changed to Gly. A round of rigid body refinement at 3.5-Å resolution was followed by a cycle of positional refinement at 2.7-Å resolution, addition of the H71 side chains to the model, and a second cycle of positional refinement. The refinement was completed with one cycle each of individual *B* value refinement with TNT and X-PLOR (Table II). Manual changes of the model, other than placement of the H71 side chain, were needed only for the H71A complex.

Refinement of the structure of the uncomplexed HuLys H71V Fv began with H71 changed to Gly. First, a round of rigid body refinement was conducted at 3.5-Å resolution. Next came two rounds of positional refinement at 2.9 Å, alternating with model-building and addition of the H71 side chains. Group *B* values (1 *B* per residue) were refined with X-PLOR, and a final cycle of positional refinement was performed (Table II).

No solvent molecules are present in any of the models. PROCHECK (25) analyses of the five structures show no residues in disallowed regions other than L51, which is in a γ -turn conformation, as seen in the native and other related structures (26).

Results

H27S structure

The structure of the HuLys Fv mutant H27S was determined as a lysozyme complex in a crystal form identical with the complex structure obtained previously (17). The crystallographic asymmetric unit contains two Fv:Ag complexes, which we designate molecule 1 and molecule 2. Both Fvs superpose well on the corresponding Fvs of the H27F structure, with root mean square (rms) differences in C α position of 0.5 Å for each of the two complexes. Despite these identical rms differences, two different conformations are present in the two crystallographically independent H27S molecules. Comparing molecule 1 of H27F and H27S, differences in C α position of up to 2.7 Å occur at residues H23–H29, adjacent to CDR-H1. The overall effect is that in the H27S structure, this portion of the molecule has moved away from the position of the

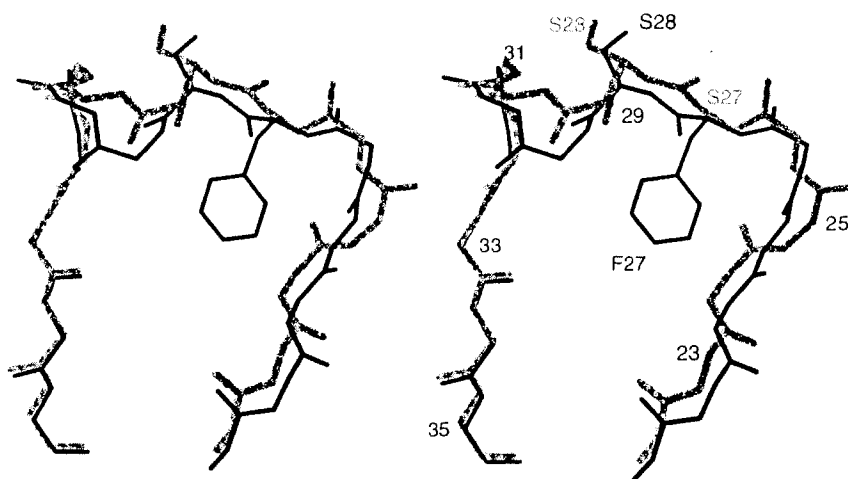
Table II. Refinement

| Parameter | H27S Complex | H71V Complex | H71A Complex | H71R Complex | H71V |
|---|--------------|--------------|--------------|--------------|----------|
| Resolution (Å) | 10.0–2.7 | 10.0–2.7 | 10.0–2.7 | 10.0–2.7 | 10.0–2.9 |
| Reflections | | | | | |
| Total (<i>F</i> >2 σ) | 20,005 | 19,501 | 18,592 | 19,535 | 14,455 |
| Working set | 18,105 | 17,642 | 16,805 | 17,653 | 13,014 |
| Test set | 1,900 | 1,859 | 1,787 | 1,882 | 1,441 |
| Atoms | 5,478 | 5,486 | 5,482 | 5,494 | 3,484 |
| <i>R</i> value ^a | | | | | |
| Working | 0.203 | 0.202 | 0.202 | 0.207 | 0.225 |
| Free | 0.313 | 0.291 | 0.291 | 0.297 | 0.279 |
| rms deviation from ideality | | | | | |
| Bond lengths (Å) | 0.015 | 0.014 | 0.015 | 0.014 | 0.021 |
| Bond angles (°) | 1.8 | 1.8 | 1.9 | 1.8 | 2.5 |
| PROCHECK analysis | | | | | |
| % in most favored regions | 80.9 | 81.5 | 80.5 | 78.8 | 78.6 |
| Estimated error in atomic position (Å) ^b | 0.33 | 0.34 | 0.33 | 0.34 | 0.37 |

$$^a R = \sum_{hkl} |F_{obs,hkl} - F_{calc,hkl}| / \sum_{hkl} |F_{obs,hkl}|$$

^b Calculated by method of Luzzati (27).

FIGURE 1. Structural effects of different amino acids at position H27, molecule 1. Stereo view of H chain CDR 1 and adjacent peptide segment in H27F (black) and H27S (gray). Atomic coordinates were taken from molecule 1 from the H27F and H27S Fv-lysozyme complex structures. The entire H27S-lysozyme complex was superposed on the H27F-lysozyme complex. The superposed molecules were used for this illustration. Only the peptide backbone from residues H22 to H35 and the side chains from residues H27 and H28 were drawn, to make clear the conformational changes that occur when Ser or Phe is substituted at position H27. Residue H27 in the H27S mutant was modeled as Gly.



Phe side chain present in H27F, toward the H chain N terminus and lysozyme, creating a more open loop (Fig. 1). Residues H74–H76, which pack against CDR-H1, have moved into the space created by this shift. Eight of the C α shifts larger than twice the rms difference come from residues H23–H29 and H74 and H76. (The others are at chain termini or at locations remote from the combining site.) Modeling of residues H22–H31 was difficult, and the side chain at position H27 could not be fit at all. The possibility exists that this remodeled region is in more than one conformation.

H27S molecule 2 shows a clear difference from the corresponding molecule 2 of the H27F Fv. The Ser and Phe side chains at the substitution site point in opposite directions. As evident in Fig. 2, the phenyl ring of H27F is buried in the interior of the extended loop formed by residues H23–H35, whereas the Ser side chain in H27S points to the aqueous exterior. As predicted (4, 9), substitution of Ser for Phe has created a cavity. Residue Ser H28 in the H27S Fv has shifted so that its main chain and side chain have moved into space occupied by the Phe H27 side chain in the H27F Fv. This large perturbation in backbone conformation extends for several residue positions along the peptide backbone, as is evident in Fig. 2. The shifts of the C α atoms of residues H23–H31 account for 9 of the 13 shifts greater than twice the rms difference between the mutant and native complexes. A shift at H75 accounts for one more, and the others are at chain termini.

Although the conformation of the loop preceding CDR-H1 differs significantly in H27S and H27F, structural effects on lysozyme binding are small. In the D1.3 complex structure, residue H32 of CDR-H1 makes a weak (3.5 Å) direct contact with lysozyme. Residues H30 and H31 make contact via water molecules (13). In the HuLys H27S structure, the distance for the potential direct contact between H32 and lysozyme is 4.1 Å (molecule 1) or 4.3 Å (molecule 2), similar to the 4.0-Å contact seen in the H27F molecule 1 complex and an increase from the 3.4 Å in the H27F molecule 2 complex, and too large to be important in lysozyme binding (28). Due to the resolution of x-ray data for the HuLys complexes, we have not modeled water molecules, hence we cannot directly compare the Fv-lysozyme interactions involving residues H30 and H31 to the corresponding interactions in D1.3. However, we did compare the positions of the Fv atoms in H27S and H27F involved in these contacts, the carbonyl oxygen atoms of H30 and H31. Both these atoms in H27S molecule 1 have moved 0.8 Å from their positions in H27F. In H27S molecule 2, the backbone C α atoms of these residues have moved 1.9 Å (H30) and 1.2 Å (H31) from their positions in the H27F complex. The atoms actually forming the contacts, H30 O and H31 O, have moved 1.7 Å and 0.9 Å, respectively. The size of this shift does not necessarily mean that these contacts are broken. The water molecules in the H27S complex presumably could shift position to accommodate the new

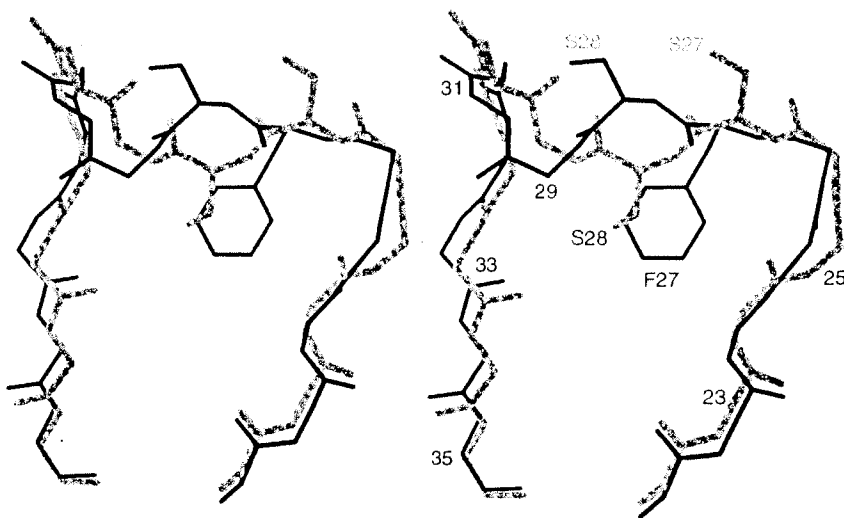


FIGURE 2. Structural effects of different amino acids at position H27, molecule 2. Stereo view of H chain CDR 1 and adjacent peptide segment in H27F (black) and H27S (gray). The illustration was prepared as for Fig. 1, but using atomic coordinates from molecule 2 of the H27F and H27S Fv-lysozyme complex structures.

Table III. *Rms changes in C α position of H71 mutants relative to H71K*

| Mutant | RMS Deviation (Å) | |
|--------|-------------------|------------|
| | Molecule 1 | Molecule 2 |
| H71V | 0.3 | 0.3 |
| H71A | 0.2 | 0.2 |
| H71R | 0.1 | 0.1 |

positions of the protein atoms. The remainder of CDR-H1 in H27S is offset from its location in H27F, with the respective chains back in register by residue H35, the last residue in the CDR.

H71 structures

The size of the side chain at position H71 is thought to control the relative disposition of loops forming CDR-H1 and CDR-H2 (29). Previously published structures of free HuLys Fv and the HuLys-lysozyme complex had Lys in this position. Here we report additional structures with Val, Ala, and Arg at residue H71. All three forms crystallized and were determined as an Fv-lysozyme complex, and a structure of the free H71V Fv was obtained as well.

All the Fv-lysozyme complexes were virtually identical. Superposition of the C α atoms of the mutant complexes onto the H71K complex gave small rms differences of 0.3 Å or less, as presented in Table III. Twelve C α atoms in the two H71V molecules have shifts greater than twice the rms differences, and none are near the combining site. Comparing the structures of the H71A and H71K complexes, four C α atoms have shifts greater than twice the rms difference; three are in the L chain and one is in lysozyme. All are remote from the combining site. The most conservative H71 substitution, arginine for lysine, gave the smallest overall rms difference. However, as for the other H71 mutants, there were moderate shifts of the mutated residue and residues in the nearby segment of polypeptide chain. The C α atoms of H71 in molecules 1 and 2 moved 0.5 Å and 0.3 Å, respectively, and the preceding C α atoms in molecule 1, H69 and H70, moved 0.2 Å and 0.4 Å. All other shifts greater than twice the rms distance occurred distant from H71 and from the combining site. Fig. 3 shows superposition of H71 and parts of CDR-H1 and CDR-H2 for the four molecules, taken from the complexed crystal forms. This illustration shows clearly that there is no change in structure of the two CDRs, despite the mutations at H71.

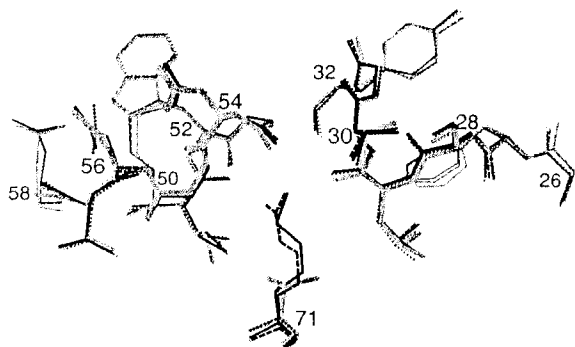


FIGURE 3. Structure of residue H71 and first and second hypervariable loops in four lysozyme-Fv complexes. Conformations of these residues in the two crystallographically independent asymmetric units of all four mutants are essentially identical. This illustration is a composite of superposed molecule 2s seen in H71K (black line), H71V (gray line), H71A (dotted line), and H71R (dashed line). Superpositions were based on H71K molecule 2 and used the C α atomic coordinates of the residues shown.

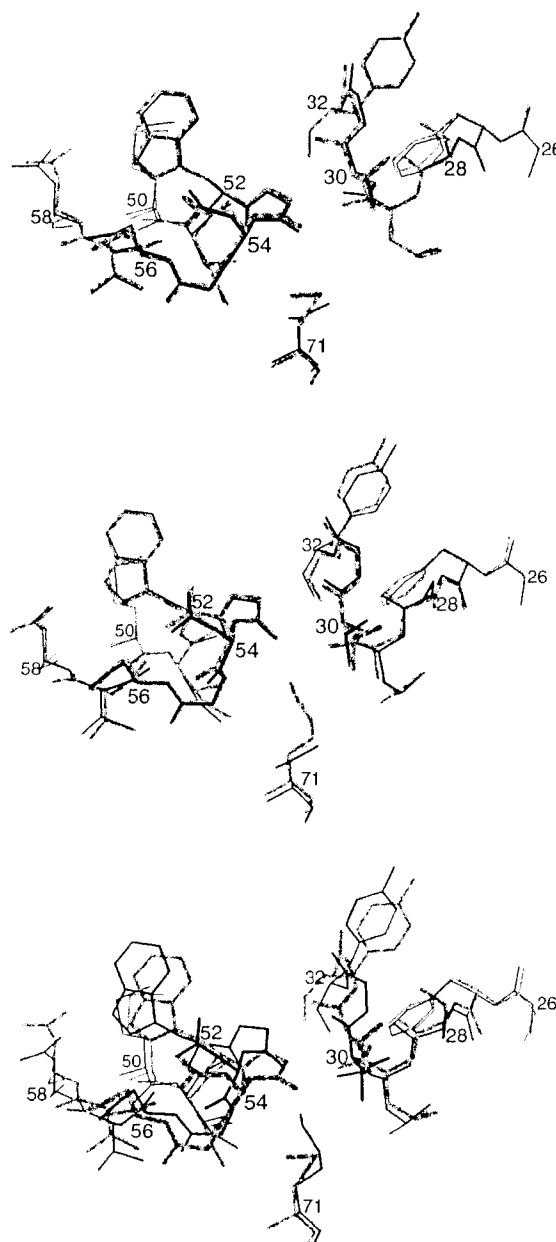


FIGURE 4. Structures of residue H71 and first and second hypervariable loops in unliganded Fv molecules. *Top*, H71V molecule (gray line) 1 superposed on H71K molecule 1 (black line). *Middle*, H71V molecule 2 (gray) superposed on H71K molecule 2 (black). *Bottom*, H71K molecule 2 (gray) superposed on H71K molecule 1 (black).

The structures of uncomplexed H71V and H71K offer another opportunity to test for a mutation-induced conformation change following the Tramontano model. The two unliganded crystal forms of H71K and H71V each have two molecules in the asymmetric unit, hence comprise a total of four independent Fv structures. Molecule 1 of H71K and molecule 1 of H71V superpose almost exactly in the region of the mutation, as evident in Fig. 4, *top*. Molecule 2 of H71K and molecule 2 of H71V superpose similarly well (Fig. 4, *middle*). However, these two pairs represent distinct conformations. The two independent molecules of H71K do not superpose well (Fig. 4, *bottom*), and the same is true for molecule 1 and molecule 2 of H71V. In other words, two Fvs

differing at residue H71, but in identical crystal packing environments, are closer in conformation than two with the same sequence, but in different environments. The two unliganded conformations observed presumably are distinct because of crystal packing interactions, rather than amino acid sequence differences at residue 71. The conformation of H71 and the two loops in the H71K and H71V Fv-lysozyme complexes is intermediate between the two conformations in the unliganded structures, though closer to molecule 2 (rms difference 0.2 Å for the C α atoms in the illustration superposed on molecule 2 of the H71K complex) than to molecule 1 (0.4 Å rms difference).

Discussion

The role of residues H26–H30 in Ag binding by humanized Abs has been ambiguous. This segment is not considered part of “Kabat” CDR-H1 (residues H31–H35), and these residues rarely contact Ag (30). Other homology- and structure-based definitions of the first Ig H chain CDR have similarly designated residues outside this segment, viz H31–H37 (31), H31–H32 (32), and H30–H35 (33). One exception is the canonical H chain hypervariable loop 1 proposed by Chothia and Lesk (34), extending from H26 to H32. The rationale for this assignment was that the segment forms a loop connecting two β -strands of rather standard geometry. The conservation of particular features in the N-terminal portion of the segment, such as an invariant Gly at position H26, was considered critical for maintaining the backbone conformation of the Ag-contacting C-terminal portion of the H26–H32 loop.

How are conformational changes in CDR-H1 transmitted from the H26–H30 region? Comparison of side-by-side crystal structures of mouse and humanized versions of the same Ab would seem a straightforward way to discover this mechanism, as identical CDR-H1 sequences are abutted in the two cases to H26–H30 regions of separate murine and human origin. However, existing structural data on humanized Abs have been equivocal.

The canonical H26–H32 structure, which the vast majority of Abs adopt (35), is typified by the human H chain NEWM (14, 34). The rat anti-CD52 Ab CAMPATH-1G, with H26–H30 sequence GFTFT, follows this canonical structure precisely (36). The initial humanized form, though based on NEWM frameworks, sequence GSTFS, bound Ag poorly, and probably did not adopt a canonical conformation. The crystallographically studied humanized form, CAMPATH-1H, had higher affinity by virtue of the H26–H30 region being reverted to the rat sequence. Nevertheless, this structure still differed from the canonical conformation at residues H29 and H30. This deviation was attributed to a different interaction with respective side chains at position H71 (Arg in CAMPATH-1G, Val in CAMPATH-1H). A recent structure of the same humanized molecule in complex with an Ag mimotope showed that the H26–H32 loop was once again in the canonical conformation (37).

CDR-H1 of the murine anti- γ -IFN Ab AF2 deviates in conformation from the canonical structure at each position, but is still topologically recognizable as a loop (38). In contrast, the humanized version of AF2, despite having an identical sequence from residue H19–H66, has an α -helical CDR-H1 not seen in any other Ab structure. This unique conformation was attributed to a second structural rearrangement in framework 1 associated with a Pro (mouse) to Ser (humanized) mutation at position H7.

The murine anti-lysozyme Ab D1.3 has a canonical CDR-H1 structure (11). The humanized version of D1.3 whose structure we previously reported (16, 17) has an identical sequence from H26–H35 (7) (H26–H30 sequence GFSLT) and also adopts a canonical CDR-H1 conformation. A kinetic study of HuLys mutants showed that a Ser substitution at residue H27 had only a slightly detrimental effect on Ag affinity (9). This observation was contrary to the

profound effect of a Ser-to-Phe mutation in CAMPATH-1H, even though both HuLys and CAMPATH-1H used NEWM framework sequences (4). One possible explanation is that the mutation in HuLys caused no significant structural change. The finding that CDR-H1 of D1.3 contributes little free energy toward lysozyme binding (39) makes plausible an alternative possibility, that the mutation did cause a change in residues H26–H30, but this perturbation was not detectable by kinetic analysis. Crystallographic data presented here favor the latter proposition, made clear in Fig. 1. The HuLys H27S structure shows large changes in backbone conformation in residues H22–H30 in molecule 1 and H26–H30 in molecule 2, but these torsional changes are not transmitted to the nearby Ag binding residues H31 and H32. Translational changes are also not transmitted to these residues, except for a displacement of H31 in molecule 2. Given our findings and the apparent idiosyncrasies observed in other humanized Ab structures, we can only conclude that the conformation of CDR-H1 and the adjacent H26–H30 region are extremely sensitive both to their own sequences and to interactions with adjacent residues. Our understanding of structural determinants of H26–H35 and our ability to rationally manipulate this region remain limited.

Tramontano et al. (29) have articulated a descriptive and predictive model for the structures of the H chain hypervariable loops 1 (Kabat residues H26–H32) and 2 (Kabat residues H52a/53–H55). In this model, the most important determinants of the conformation of hypervariable loop 2 are the length of the loop and specific sequence constraints, with particular canonical structures and conserved residues expected for 3, 4, and 6 residue loops. A further structural determinant is the side chain of residue H71, which is significant in the following way. The position of hypervariable loop 1 is essentially fixed. The position of loop 2 relative to loop 1 is variable, and depends on whether a large side chain at H71 packs between the two loops and separates them or a small side chain at H71 allows loops 1 and 2 to juxtapose.

In HuLys crystal structures with four different side chains at residue H71, the expected conformational rearrangement of the hypervariable loop 2 region is not observed. The absence of a mutation-induced conformation change cannot simply be due to the stabilizing effect of a bound Ag, because the Lys-to-Val mutation in the unliganded crystal forms also does not alter the position of loop 2. The modest (0.4–0.6 kcal/mol) improvement in affinity that accompanied this mutation thus cannot be attributed to relieving an inappropriate displacement of hypervariable loop 2 (9). Our findings do not invalidate the Tramontano model, for which other proof exists, including a specific mutational study of residue H71 in the crystallographically determined Ab B72.3 (40). Our data do demonstrate that a class of exceptions may exist in which the H71 side chain alone does not affect the separation of hypervariable loops 1 and 2. An unknown sequence determinant may override the action of H71, or the compact nature of 3-residue hypervariable loops (H53–H55) may confer less sensitivity to the bulk of the H71 side chain.

The observation that significant conformational changes in the H27S mutant did not lead to much change in Ag affinity, whereas substitutions at H71 gave affinity differences, but no apparent explicatory change in structure illustrates the value of combining structural and kinetic studies.

References

1. Wu, T. T., and E. A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132:211.
2. Kabat, E. A., T. T. Wu, H. M. Perry, K. S. Gottesman, and K. Coeller. 1991. *Sequences of Proteins of Immunological Interest*, 5th Ed. U.S. Department of

- Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD.
3. Jones, P. T., P. H. Dear, J. Foote, M. S. Neuberger, and G. Winter. 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 321:522.
 4. Riechmann, L., M. Clark, H. Waldmann, and G. Winter. 1988. Reshaping human antibodies for therapy. *Nature* 332:323.
 5. Stephens, S., S. Emtage, O. Vetterlein, L. Chaplin, C. Bebbington, A. Nesbitt, M. Sopwith, D. Athwal, C. Novak, and M. Bodmer. 1995. Comprehensive pharmacokinetics of a humanized antibody and analysis of residual anti-idiotypic responses. *Immunology* 85:668.
 6. Winter, G., and W. J. Harris. 1993. Humanized antibodies. *Immunol. Today* 14:243.
 7. Verhoeven, M. E., C. Milstein, and G. Winter. 1988. Reshaping human antibodies: grafting an antilysozyme activity. *Science* 239:1534.
 8. Riechmann, L., J. Foote, and G. Winter. 1988. Expression of an antibody Fv fragment in myeloma cells. *J. Mol. Biol.* 203:825.
 9. Foote, J., and G. Winter. 1992. Antibody residues affecting conformation of the hypervariable loops. *J. Mol. Biol.* 224:487.
 10. Mariuzza, R. A., D. L. Jankovic, G. Boulot, A. G. Amit, P. Saludjian, A. Le Guern, J. C. Mazié, and R. J. Poljak. 1983. Preliminary crystallographic study of the complex between the Fab fragment of a monoclonal anti-lysozyme antibody and its antigen. *J. Mol. Biol.* 170:1055.
 11. Amit, A. G., R. A. Mariuzza, S. E. V. Phillips, and R. J. Poljak. 1986. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 233:747.
 12. Bhat, T. N., G. A. Bentley, T. O. Fischmann, G. Boulot, and R. J. Poljak. 1990. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature* 347:483.
 13. Bhat, T. N., G. A. Bentley, G. Boulot, M. I. Green, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, and R. J. Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* 91:1089.
 14. Saul, F. A., and R. J. Poljak. 1992. Crystal structure of human immunoglobulin fragment Fab New refined at 2.0 Å resolution. *Proteins* 14:363.
 15. Epp, O., E. E. Lattman, M. Schiffer, R. Huber, and W. Palm. 1975. The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. *Biochemistry* 14:4943.
 16. Holmes, M. A., and J. Foote. 1997. Structural consequences of humanizing an antibody. *J. Immunol.* 158:2192.
 17. Holmes, M. A., T. N. Buss, and J. Foote. 1998. Conformational correction mechanisms aiding antigen recognition by a humanized antibody. *J. Exp. Med.* 187:479.
 18. Carter, P., R. F. Kelley, M. L. Rodrigues, B. Snedecor, M. Covarrubias, M. D. Velligan, W. L. T. Wong, A. M. Rowland, C. E. Kotts, M. E. Carver, et al. 1992. High level *Escherichia coli* expression and production of a bivalent humanized antibody fragment. *BioTechnology (NY)* 10:163.
 19. Perkins, S. J. 1986. Protein volumes and hydration effects. *Eur. J. Biochem.* 157:169.
 20. Otwinowski, Z. 1993. Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing, January 29–30*. L. Sawyer, N. Isaacs, and S. Bailey, eds. SERC Daresbury Laboratory, Warrington, U.K. pp. 56.
 21. Otwinowski, Z., and W. Minor. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307.
 22. Kleywegt, G. J., and A. T. Brünger. 1996. Checking your imagination: applications of the free R value. *Structure* 4:897.
 23. Brünger, A. T. 1992. *X-PLOR Manual, Version 3.1*. Yale University Press, New Haven, CT.
 24. Tronrud, D. E., L. F. TenEyck, and B. W. Matthews. 1987. An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. A* 42:489.
 25. Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* 26:283.
 26. Milner-White, E. J., B. M. Ross, R. Ismail, K. Belhadj-Mostefa, and R. Poet. 1988. One type of γ -turn, rather than the other gives rise to chain-reversal in proteins. *J. Mol. Biol.* 204:777.
 27. Luzzati, V. 1952. Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* 5:802.
 28. Baker, E. N., and R. E. Hubbard. 1984. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* 44:97.
 29. Tramontano, A., C. Chothia, and A. M. Lesk. 1990. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the V_H domains of immunoglobulins. *J. Mol. Biol.* 215:175.
 30. Padlan, E. O., C. Abergel, and J. P. Tipper. 1995. Identification of specificity-determining residues in antibodies. *FASEB J.* 9:133.
 31. Capra, J. D. 1971. Hypervariable region of human immunoglobulin heavy chains. *Nat. New Biol.* 230:61.
 32. Novotny, J., R. Brucoleri, J. Newell, D. Murphy, E. Haber, and M. Karplus. 1983. Molecular anatomy of the antibody binding site. *J. Biol. Chem.* 258:14433.
 33. MacCallum, R. M., A. C. R. Martin, and J. M. Thornton. 1996. Antibody-antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.* 262:732.
 34. Chothia, C., and A. M. Lesk. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* 96:901.
 35. Chothia, C., A. M. Lesk, E. Gherardi, I. M. Tomlinson, G. Walter, J. D. Marks, M. B. Llewellyn, and G. Winter. 1992. Structural repertoire of the human V_H segments. *J. Mol. Biol.* 227:799.
 36. Cheetham, G. M. T., G. Hale, H. Waldmann, and A. C. Bloomer. 1998. Crystal structures of a rat anti-CD52 (CAMPATH-1) therapeutic antibody Fab fragment and its humanized counterpart. *J. Mol. Biol.* 284:85.
 37. James, L. C., G. Hale, H. Waldmann, and A. C. Bloomer. 1999. 1.9 Å structure of the therapeutic antibody CAMPATH-1H Fab in complex with a synthetic peptide antigen. *J. Mol. Biol.* 289:293.
 38. Fan, Z. C., L. Shan, B. Z. Goldstein, L. W. Guddat, A. Thakur, N. F. Landolfi, M. S. Co, M. Vasquez, C. Queen, P. A. Ramsland, and A. B. Edmundson. 1999. Comparison of the three-dimensional structures of a humanized and a chimeric Fab of an anti- γ -interferon antibody. *J. Mol. Recognit.* 12:19.
 39. Dall'Acqua, W., E. R. Goldman, E. Eisenstein, and R. A. Mariuzza. 1996. A mutational analysis of the binding of two different proteins to the same antibody. *Biochemistry* 35:9667.
 40. Xiang, J., Y. Sha, Z. Jia, L. Prasad, and L. T. J. Delbaere. 1995. Framework residues 71 and 93 of the chimeric B72.3 antibody are major determinants of the conformation of heavy-chain hypervariable loops. *J. Mol. Biol.* 253:385.

Somatic Hypermutation Introduces Insertions and Deletions into Immunoglobulin V Genes

By Patrick C. Wilson,^{*†} Odette de Bouteiller,[§] Yong-Jun Liu,[§]
Kathleen Potter,^{*} Jacques Banchereau,^{||} J. Donald Capra,^{*†}
and Virginia Pascual^{*}

From the ^{}Molecular Immunology Center, Department of Microbiology, University of Texas Southwestern Medical Center at Dallas, Texas 75235-9140; the [†]Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104; [§]Schering-Plough Laboratory for Immunological Research, Dardilly, France; and ^{||}Baylor Institute for Immunological Research, Dallas, Texas 75235*

Summary

During a germinal center reaction, random mutations are introduced into immunoglobulin V genes to increase the affinity of antibody molecules and to further diversify the B cell repertoire. Antigen-directed selection of B cell clones that generate high affinity surface Ig results in the affinity maturation of the antibody response. The mutations of Ig genes are typically base-pair substitutions, although DNA insertions and deletions have been reported to occur at a low frequency. In this study, we describe five insertion and four deletion events in otherwise somatically mutated V_H gene cDNA molecules. Two of these insertions and all four deletions were obtained through the sequencing of 395 cDNA clones (~110,000 nucleotides) from CD38⁺IgD⁻ germinal center, and CD38⁻IgD⁻ memory B cell populations from a single human tonsil. No germline genes that could have encoded these six cDNA clones were found after an extensive characterization of the genomic V_H4 repertoire of the tonsil donor. These six insertions or deletions and three additional insertion events isolated from other sources occurred as triplets or multiples thereof, leaving the transcripts in frame. Additionally, 8 of 9 of these events occurred in the CDR1 or CDR2, following a pattern consistent with selection, and making it unlikely that these events were artifacts of the experimental system. The lack of similar instances in unmutated IgD⁺CD38⁻ follicular mantle cDNA clones statistically associates these events to the somatic hypermutation process ($P = 0.014$). Close scrutiny of the 9 insertion/deletion events reported here, and of 25 additional insertions or deletions collected from the literature, suggest that secondary structural elements in the DNA sequences capable of producing loop intermediates may be a prerequisite in most instances. Furthermore, these events most frequently involve sequence motifs resembling known intrinsic hotspots of somatic hypermutation. These insertion/deletion events are consistent with models of somatic hypermutation involving an unstable polymerase enzyme complex lacking proofreading capabilities, and suggest a downregulation or alteration of DNA repair at the V locus during the hypermutation process.

During the course of a T cell-dependent antibody response, B cells hone the specificity of their antibody molecules through a process of random somatic hypermutation of their V genes, followed by antigen driven selection. This is collectively referred to as affinity maturation. This process occurs within the germinal centers (GCs)¹ of secondary follicles from peripheral lymphoid organs when

antigen stimulated B cells receive proper signals from T and accessory cells. In the human system, GC B cells are characterized by the surface expression of CD38 and, in most cases, the loss of IgD (1–3). We have previously shown that the initiation of somatic hypermutation occurs within the CD77⁺ subset of these IgD⁻CD38⁺ B cells (4). Mutated V genes can be isolated from all subsequent stages of B cell differentiation and in cells from all IgD⁻ and certain IgD⁺ B cell subsets (4, 5). The molecular process of somatic hypermutation remains elusive, primarily due to the lack of a good in vitro model until very recently (6). Much of what

¹Abbreviations used in this paper: FM, follicular mantle; FW, framework; GC, germinal center.

is known concerns: (a) localizing the somatic hypermutation process to particular B cell subsets and anatomical settings (4, 7–10); (b) delineating the limits and rates of mutational activity (11); (c) determining the minimal substrate through transgenic technology (12, 13); and (d) analyzing the mutations themselves in the context of the surrounding sequence to reveal tendencies such as strand polarity and “hotspots” of somatic hypermutation (for reviews see references 12 and 13).

Although somatic hypermutation is typically described as the generation of bp substitutions, insertions and deletions have been sporadically described. As with somatic point mutations, the analysis of these events can provide valuable information concerning somatic hypermutation itself. Analysis of human V_H4 family genes generated from the amplification of cDNA from somatically mutated GC (IgD^+CD38^+) and memory (IgD^+CD38^-) B cell subpopulations led us to identify a number of cDNA clones from the mutated cell populations that contained insertions and deletions. We provide evidence that these events are linked to the somatic hypermutation process. Additionally, these events occur in a predictable fashion relative to the surrounding sequence, suggesting a model for their occurrence with implications for the molecular process of somatic hypermutation.

Materials and Methods

Isolation, Labeling, and Sorting of Tonsil B Cells. Human tonsils were obtained during routine tonsillectomy. B cell isolation and sorting for CD38 and IgD expression were performed as previously described (4, 14). In brief, human tonsillar B cells were separated into IgD^+CD38^- follicular mantle (FM) B cells, IgD^+CD38^+ GC B cells, and IgD^+CD38^- memory B cells to 95–98% purity as predicted by FACS[®] analysis, as previously described (13). The mutation state of the V_H gene cDNA clones from the various subpopulations was in agreement with our previous study (4). Clones were considered somatically mutated if they contained two or more bp substitutions, well beyond the expected error rates for the avian myeloblastosis virus reverse transcriptase (AMV-RT), Taq, and PFU polymerases used in these analyses (this mutation rate is based on our previous analyses; reference 4).

Sequencing the $Ig V_H$ Transcripts. Total RNA was extracted from $1-5 \times 10^5$ B cells using guanidinium thiocyanate-phenol-chloroform in a single step using the Ultraspec RNA isolation system (BIOTECX Laboratories, Houston, TX), and was reverse transcribed using oligo-d(T) or specific V gene constant region oligonucleotides $C_{\mu}12$ (5'-CTGGACTTTGCACACCACGTG-3') for IgM transcripts or $C_{\gamma}180$ (5'-CTGCTGAGGAGTAGACTCC-3') for IgG transcripts and SuperScript II reverse transcriptase (GIBCO BRL, Gaithersburg, MD). First strand cDNA was used directly for second strand synthesis and amplification via PCR using internal primers corresponding to the C_{μ} or C_{γ} constant regions in combination with V_H4 or V_H6 family-specific leader oligonucleotides: $C_{\gamma}140$, 5'-GGCAAGGTGTGCACGCCGCTG-3'; $C_{\mu}10$, 5'-TCTGTGCC CTGCATGACGTC-3'; L-4, 5'-ATGAAACACCTGTGTTCTT-3'; L-6, 5'-ATGTCTGTCTCCTTCCTCAT-3'. The PCR products were purified using microconcentrators (Amicon, Beverly, MA), and then were kinased and blunt-end ligated into an EcoRV-digested and dephos-

phorylated pBluescript plasmid (Stratagene, La Jolla, CA; Polynucleotide Kinase, T4 DNA Ligase, and EcoRV were from Boehringer Mannheim, Amsterdam, Netherlands). After transformation by electroporation into electro-competent DH10 α *Escherichia coli* (GIBCO BRL) and screening with consensus internal oligonucleotides as previously described (4, 15), positive colonies were picked, plasmid mini-preparations were made, and colonies were sequenced in both directions using an automated DNA sequencer and automated sequencer protocol (ABI-377; Advanced Biotechnologies Inc., Columbia, MD). All sequences were analyzed using DNASTar (DNASTar Inc., Madison, WI). In the first tonsil analyzed, 583 clones were picked, plasmid mini-preparations were made, and Southern blots were prepared by standard methods. These blots were screened with a set of oligonucleotides specific for the various V_H4 family genes. Only those clones that screened positive with constant region probes but negative for the various V_H4 complementarity-determining region (CDR)1-specific probes were sequenced (395 of 583 clones), thus enriching the somatically mutated populations analyzed, in that the CDR1 probes should anneal only to the sequences most similar to germline. The frequency of the occurrence of these events can therefore only be predicted to be between 6 out of 395 and 6 out of 583 clones (1–2%). Any sequence of interest was resequenced in both directions to ensure sequence fidelity.

Characterizing the Genomic Repertoire. Total genomic DNA was isolated from FM B cells (IgD^+ , $CD38^-$) using the Puregene DNA isolation kit (Gentra Systems, Inc., Minneapolis, MN). V_H4 genes were amplified using a V_H4 leader-specific primer (L-4, as above) and a primer specific for all V_H4 gene family heptamer-nonamer spacer regions as previously described (16). PCR products were agarose gel purified, then cloned into *E. coli* as described above for the cDNA clones. Clones identified in the cDNA analysis that contained insertions or deletions were used to design PCR primers to amplify both the exact sequence of clones with insertions/deletions as found and the predicted sequences based on the proposed germline counterparts. Oligonucleotides used in this analysis (Format, is as follows: clone: exact/predicted): g64:5'-GGACGGTGTGTAAGTGGTTCC-3'/5'-GGACGGGTTGTAGGTC-TCC-3'; g144:5'-TCTTGAGGGACGGGTTGGTGT-3'/5'-TCTTGAGGGACGGGTTGT-3'; g187:5'-CAGCTCCAGTAGTAAGCCCCG-3'/5'-CAGCTCCAGTAGTAACCACCG; g188:5'-GAGGGATTGTAGTTGGAGCC-3'/5'-GAGGGGTTGTAGTTGGTCCC; g192:5'-CCAGCCCCAGTAGTAGTAAGT-3'/5'-CCAGCCCCAGTAGTAGTAAGT-3'; g80:5'-GCGGATCCAGTAGTAACC-3'/5'-GCGGATCCAGTAGTAACC-3'.

Sequence Availability. All cDNA sequences with insertions or deletions, and any genomic sequences unique to the literature as described in the results section are available from EMBL/Genbank/DBJ under accession numbers AF013615 through AF013626.

Assay for Screening V_H Gene Lengths. To facilitate the analysis of large numbers of V_H gene transcripts for the presence of insertions or deletions, first strand cDNA produced as described above was PCR amplified using Expand high fidelity polymerase (Boehringer Mannheim) to reduce errors resulting from Taq polymerase alone. The products of this PCR amplification were cloned as described above and screened using ³²P-labeled, gene-specific oligonucleotides (V_H4 -39:5'-ATTGGGAGTATCTATTATAGT-3'; L-6 as above). Positive colonies were picked and used to inoculate overnight cultures. A 1 μ l aliquot from each 24-h culture was used to directly inoculate 25- μ l PCR amplification mixtures in 96-well-format PCRs. The internal PCR reactions used ³²P-labeled, gene-specific oligonucleotides to amplify a 230-base fragment including the V_H4 -39 CDR1 (L-4, as above, and V_H4 -39-3': 5'-

A) Insertion events from a single tonsil:

4-39 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q144 -----R-----G-----I-S-----L-S-----T-----N-----IY-----

4-39 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q192 -----K-----R-----A-----A-----Y-----V-----V-----G-----A-----R-----N-----F-----

B) Deletion events from a single tonsil:

4-59 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q188 -----I-----L-S-----V-----R-S-----S-----E-----T-----R-----S-----

4-31 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q187 -----R-----S-----AFL-----A-----A-T-C-----I-----N-L-V-LY-NH-----F-R-G-----

4-31 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q64 -----H-----R-----A-----A-----F-----TK-----S-----N-T-----ACPF-----

4-34 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
q80 -----S-----SK-----VR-----R-----I-----L-F-R-----T-----I-----

C) Other insertions:

V_H6 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
h8p2 -----S-GV-----

V_H4-31 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
pg86 -----P-----D-----FTSGDSF-----F-----L-----L-----H-D-----G-D-----L-MTI-----L-----A-L-----

V_H6 QVLQESGPGLVKPSSETLSLTCTVSGGSISSSYTWGWIROPFGKLEWIGSIYYSGSTYY NPSLKSAVTISVDTSKNQFSLKLSVTAADTAVYYCAR
tm121 -----R-----R-----D-----

Figure 1. Predicted amino acid sequences of nine cDNA clones with insertions or deletions. Note that all of these clones are extensively mutated. In all but clone tm121, the insertions or deletions occur in CDR1 or CDR2. (A and B) Two clones with insertions (A) and four clones with deletions (B) from a single tonsil. (C) Three additional cDNA clones with insertions isolated from various sources. Sequence data available from GenBank/DBJ under accession numbers AF013615 through AF013626.

GCTCCCACTATAATAGATACT-3') or for analysis of V_H6 genes a 166-nucleotide fragment including the CDR1 and CDR2 of V_H6 (V_H6FW1: 5'-TGCCATCTCCGGGGACAGTGT-3', V_H6FW3: 5'-TGTGTCTGGGTTGATGTTAT-3'). Aliquots of each clone were also used to inoculate amplifications of CDR3 regions using FW3-specific (ssFW3: 5'-CTGAA[C/G]CT-GAGCTCTGTGAC[T/C]) and C_μ- or C_γ-specific oligonucleotides (C_μD: 5'-GGAATTCTCACAGGAGACGA-3', C_γ-140 as above) to analyze the diversity of the populations under study; the distribution of CDR3 size variations of several hundred V_H sequences cloned in this analysis were used to produce an expected distribution of CDR3 sizes for comparison (see Fig. 5 B). The amplification products were electrophoresed on 0.6X-TBE, 5% urea-acrylamide sequencing gels (Long Ranger; J.T. Baker, Phillipsburg, NJ) and analyzed with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA.) using the Image Quant software supplied by the manufacturer. Clones that differed from the expected size and those clones in lanes adjacent to aberrantly migrated bands were used to produce plasmid preparations from which the inserts were sequenced in either direction.

Scoring of Insertion/Deletion Events. In the results section, insertions and deletions are scored as events per 10⁴ nucleotides within the customary boundaries of CDR1 and CDR2. This unit was chosen because in the selected populations studied these events are generally only found in the CDR regions and therefore the comparison of events per total nucleotides would be misleading. In the PAGE analysis, each V_H4-39 FM clone included only the CDR1 (21 nucleotides) within a total of 230 nucleotides/clone, whereas each V_H6 FM clone was only 166 nucleotides but included both the CDR1 and CDR2 (75 CDR nucleotides). In the sequencing analysis, various B cell populations were analyzed involving a wide range of overall lengths. Comparisons of the frequency of insertions/deletions just within the CDRs allowed for a more standardized and quantitative analysis, and for more freedom in experimental design.

Baculovirus Expression System. Cloning and coexpression of clone pg86 and κ light chain FS6κ in the baculovirus expression

system was performed as previously described (17). Recombinant *Autographa californica* nuclear polyhedrosis virus (AcMNPV) was cloned using the pH360NX transfer vector and expressed in Sf9 cells.

Capture ELISA for γ Heavy Chain, and κ Light Chains. Expression of recombinant antibodies of clone pg86 coexpressed with κ light chain FS6κ were measured by capture ELISA. Wells were coated with goat anti-human IgG and incubated with supernatant of recombinant pg86/FS6κ added in serial twofold dilutions. Bound antibody was detected using alkaline phosphatase-conjugated goat anti-human IgG, or goat anti-human Cκ. After 1-h incubation at 37°C, phosphatase substrate was added and absorbance was measured at 405 nm in an ELISA plate reader.

Results

Insertions and Deletions into Immunoglobulin V_H Genes. In a large scale analysis of V_H genes from both the IgM and IgG compartments of B cell subpopulations separated from a single human tonsil, six clones that contained DNA insertions or deletions were isolated. These insertions and deletions were apparently selected in that they involved nucleotide triplets or multiples of nucleotide triplets, leaving the cDNAs (transcripts) in frame, and they were localized to the CDR1 and CDR2 (Fig. 1, A and B). The six clones with insertions or deletions were identified from the sequencing of 395 cDNA clones (~110,000 nucleotides) from GC and memory B cell subpopulations, resulting in a frequency of <2% of clones analyzed (~1 event/18,000 nucleotides). All six events were in IgG transcripts. Two events were obtained from IgD⁺CD38⁺ GC and four events from IgD⁺CD38⁺ memory cell populations. None of the IgM V_H cDNAs analyzed from this tonsil had insertions or deletions, although we have observed such events in IgM transcripts in the past and in subsequent analyses, as described below.

V_H4 Gene family:

| <i>V_H4 gene*</i> | GGC | TCC | GTC | AGC | ← Customary CDR1 Boundaries → | AGT | GGT | AGT | TAC | TAC | TGG | AGC | TGG | ATC |
|-----------------------------|------|-----|-----|-----|-------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 4-61 | --- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| none | --- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4-31 | --- | --- | A-- | --- | --- | --- | --- | G-- | --- | --- | --- | --- | --- | --- |
| 4-39 | --- | --- | A-- | --- | --- | A-- | --- | --- | --- | --- | --- | G-- | --- | --- |
| 4-04B | --- | --- | A-- | --- | --- | A-- | --- | --- | GG | --- | --- | --- | --- | G-- |
| 4-11 | --- | --- | A-- | --- | --- | A-- | --- | --- | OG | --- | --- | --- | --- | G-- |
| 4-55 | A-- | --- | A-- | --- | --- | A-- | --- | --- | OG | --- | --- | --- | --- | G-- |
| (pseudo) | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4-28 | TA-- | --- | A-- | --- | --- | A-- | --- | --- | OG | --- | --- | G-- | --- | --- |
| 4-22 | TA-- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | G-- | --- | --- |
| 4-04 | --- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4-59 | --- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (4-59 alleles) | --- | --- | A-- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4-34 | G-- | --- | T-- | --- | --- | G-- | --- | --- | --- | --- | --- | --- | --- | --- |

*Nomenclature based on Matsuda and Honjo(37)

Figure 2. Comparison of the CDR1s of the human *V_H4* germline genes. The primary variability between *V_H4* family members is 3–6-bp size variances in the CDR1s which is similar to the short insertions and deletions that we attribute to somatic hypermutation in the selected B cell populations studied in this report.

The Insertions and Deletions Are Not Germline Encoded. The analysis described above focused on the *V_H4* gene family, which consists of 10–14 members/genome, varying slightly between individuals (16, 18). As shown in Fig. 2, the major difference between *V_H4* genes involves the length of CDR1. Because genomic diversity between *V_H4* family members resembles the events described in this paper we had to rule out possible alternative explanations for these events, such as: (a) different alleles of the detected genes; (b) rarely expressed or otherwise unknown *V_H4* gene family members; or (c) hybrids between known and detected *V_H* genes and/or other artifacts of the experimental system. To address these issues, both the expressed and genomic repertoires from this tonsil were characterized. As indicated in Table 1, 2 out of 118 *V_H4*–39, 2 out of 49 *V_H4*–31, 1 out of 87 *V_H4*–34, and 1 out of 45 *V_H4*–59 cDNA clones contained insertion/deletion events. cDNA clones were judged as unique isolates based on CDR3 analysis, and the few isolates that appeared to be clonally related differed in their patterns of somatic mutation beyond the level explainable by reverse transcription and PCR errors (maximum: >1 mutation/500 nucleotides of *V_H* gene sequence as previously described [4]).

To characterize the genomic repertoire of the initial tonsil, 80 germline *V_H4* gene clones were isolated and sequenced (Table 1), which encompassed all 14 *V_H4* family members or alternate alleles represented in the 446 cDNA clones analyzed from all of the tonsillar B cell subsets. In the course of this study, we isolated the germline counterpart of a novel *V_H4* gene segment for which transcripts had been found. In addition, germline genes corresponding to two apparently functional *V_H4* genes not found as cDNA clones in this analysis were isolated, as well as one nonfunctional *V_H4* gene and a divergent polymorphism of a known *V_H4* pseudogene. The proposed germline counterparts of each of the *V_H4* genes containing insertion/deletion events were isolated from 4 to 11 times (Table 1). 8 independent genomic isolates of *V_H4*–31 and of *V_H4*–39 were cloned. *V_H4*–34 and *V_H4*–59 were isolated 11 and 4 times, respectively. No germline genes were isolated that could have encoded the insertion/deletion events described.

To further be certain that the insertion/deletion events

Table 1. *cDNA and Germline Clones Isolated*

| <i>V_H4 gene alleles isolated*</i> | cDNA clones with cDNA ins/del | Total clones isolated | Germline clones isolated [†] |
|---|-------------------------------|-----------------------|---------------------------------------|
| <i>V_H4</i> –39 | 2 | 113 | 7 |
| <i>V_H4</i> –31 | 2 | 49 | 8 |
| <i>V_H4</i> –59 | 1 | 45 | 4 |
| <i>V_H4</i> –34 | 1 | 87 | 11 |
| <i>V_H4</i> –34 related | 0 | 0 | 4 |
| <i>V_H4</i> –55 pseudogene [‡] | 0 | 0 | 12 |
| <i>V_H4</i> –55-related pseudogene [‡] | 0 | 0 | 3 |
| <i>V_H4</i> –04 | 0 | 17 | 7 |
| <i>V_H4</i> –04-related pseudogene [‡] | 0 | 0 | 2 |
| <i>V_H4</i> –61 | 0 | 25 | 7 |
| New <i>V_H4</i> gene [§] | 0 | 33 | 3 |
| <i>V_H4</i> –04B | 0 | 72 | 1 |
| <i>V_H4</i> –28 | 0 | 0 | 1 |

*Nomenclature based on Matsuda and Honjo (37).

[†]Nine unusual isolates were also cloned consisting of hybrids of two of the indicated genes, presumably do to PCR artifact. None of these artifacts were altered in size or resembled any of the insertion or deletion events observed.

[‡]Pseudogenes contain stop codons or frameshift mutations and are not expressed.

[§]Newly identified *V_H4* gene is most closely related to *V_H4*–04.

described herein were not germline encoded, two sets of PCR primers were designed to specifically recognize: (a) the exact sequence of the events; (b) the predicted, unmutated, germline sequence corresponding to the cDNAs containing insertion and deletion events. These primers were used to amplify genomic DNA from this individual, yielding negative results (data not shown). The unique nature of these events relative to both the expressed and genomic repertoire and our inability to amplify genomic counterparts for these events by PCR suggest that they are not germline encoded.

The Proposed Insertion/Deletion Events Are Not the Result of (V_H/V_H) Recombination. As in most V gene repertoire analyses, we detected hybrid *V_H* sequences that could be the result of either PCR splicing by overlap extension artifacts, or reciprocal homologous recombination between un-rearranged V genes (19). However, none of these likely artifactual events were altered in size such that they resembled the insertion or deletion of DNA described above. A number of artifacts of this type had been isolated in the cDNA analysis as well; such artifacts are common to V gene analyses (20). The cDNA isolates with deletion and insertion events were stringently compared to all germline and cDNA isolates and were found to be unique relative to both the expressed and germline *V_H4* gene repertoires of this individual, supporting a somatic origin for their occurrence.

The Insertions and Deletions Are Associated with Somatic Hypermutation. To determine whether or not these inser-

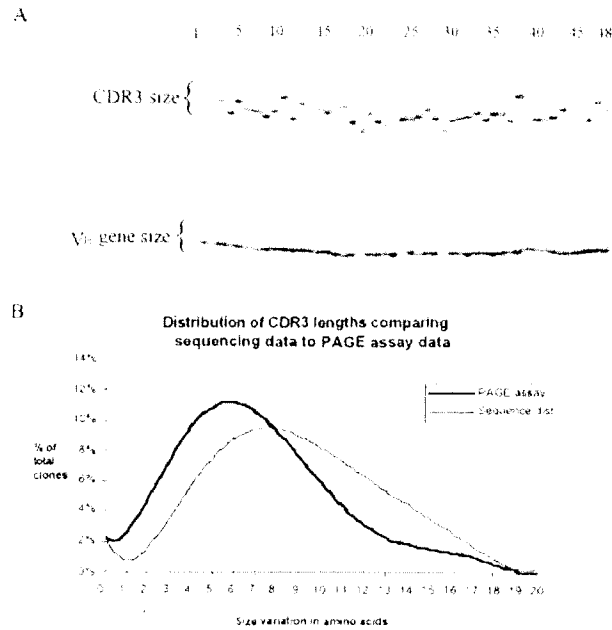


Figure 3. Polyacrylamide gel assay to identify insertions or deletions into V_H genes. (A) Phosphorimage of a polyacrylamide gel: each lane contains the hot-PCR products (^{32}P -labeled) of the V_H gene and the CDR3 of an individual clone. (B) A comparison of the distribution of CDR3 sizes of the 485 CDR3s assayed to the distribution of 500 CDR3s observed in sequences from this report indicates that the clones assayed by electrophoresis were a polyclonal population. CDR3 sizes were measured from the most 3' Tyr residue (common to all V genes analyzed) to the most 5' C μ or C γ residue. CDR3 lengths for those assayed by electrophoresis were extrapolated based on sequencing of 75 out of the 485 clones assayed. The x-axis is the number of amino acids greater than the shortest CDR3 found.

tion/deletion events were associated with somatic hypermutation, we analyzed their occurrence in unmutated FM transcripts. This was done using either direct sequencing or PCR amplification of portions of the V_H genes spanning the CDRs, followed by size comparisons on polyacrylamide gels (Fig. 3). Any clones that ran aberrantly, and the clones in adjacent lanes, were sequenced (75 out of the 485 clones). None of these 75 clones were related based on CDR3 homology. To ensure that the remaining 410 FM clones were polyclonal, the CDR3s were PCR amplified and loaded on the sequencing gels simultaneously to the V_H gene amplification products for size comparisons (Fig. 3 A). The size distribution of these CDR3s was similar to that of ~ 500 V_H gene sequences analyzed in this study (Fig. 3 B), providing evidence that our FM sample is polyclonal.

The six events detected from a single tonsil were isolated from 395 mutated cDNA clones (25,482 CDR nucleotides), corresponding to a frequency of 2.35 events/ 10^4 CDR nucleotides. This is significantly different ($p=0.014$ by a one-sided χ^2 test) from the analysis of unmutated FM-derived clones (25,515 CDR nucleotides) that yielded no insertions or deletions (Table 2).

In the course of the analysis described above, we isolated one IgM clone containing a 6-nucleotide insertion into framework (FW)3 (see below). We believe that this clone is part of the mutated GC or memory repertoire because it contained 4 bp substitutions in addition to the insertion. In this study, the B cell populations analyzed were 95–98% pure, and the FM B cell subpopulation could therefore include between 2 and 5% contaminating clones, that is, IgM-expressing cells not from the naïve population that can therefore be somatically mutated. However, none of

Table 2. Analysis of Unmutated FM cDNA Clones for Insertion or Deletion Events

| Clone type | Clones assayed | CDR nucleotides [†] | Events observed | Frequency [§] | Expected (events/ 10^4 CDR nucleotides) |
|---|----------------|------------------------------|--------------------------|------------------------|---|
| Mutated V_H4 clones (GC and memory B cells) | 395 | 25,482 | 6 | 2.35 events/ 10^4 mn | |
| Unmutated clones: | | | | | |
| V_H4 -FM, CDR1* | 265 | 5,565 | 0 | 0 | 1.31 |
| V_H6 IgM FM V_H genes* | 220 | 16,500 | 0 | 0 | 3.88 |
| V_H4 family FM sequences | 51 | 3,450 | 0 | 0 | 0.81 |
| Total unmutated values | | 25,515 | 0 | 0 | 2.35 events/ 10^4 CDR nucleotides |
| | | | $(P = 0.014)^{\ddagger}$ | | |

*Clones analyzed by hot-PCR/PAGE assay as described in the text.

[†]CDR nucleotides are those within the customary bounds of the CDR1 and CDR2. (See Materials and Methods for a more detailed explanation of this unit).

[§]Events per 10^4 CDR nucleotides.

^{||}Expected frequency (events/ 10^4 CDR nucleotides) derived from sequencing data: 6 events in 25,482 CDR nucleotides; $6/(25,482 \text{ CDR nucleotides}/10^4) = 2.35$.

[‡]Statistical analysis: χ^2 test for independence.

A Insertion events:

V_H6: GTC TCT **AGC** --- AAC AGT GCT
hBp2: GTC TCT **AGC AGC** AAC AGT GCT

V_H4-39: TAC TAC **AAC** --- CCG TCC CTC
g144: TAC TAC **AAC AAC** CCG TCC CTC

V_H4-39: AGT **TAC TAC** --- TGG GGC TGG
g192: AGT **TAC TAC TAC** TGG GGC TGG

V_H6: TCC **AAG AAC** --- --- CAG TTC
tm121: TCC **AAG AAC AAG AAC** CAG TTC

B Deletion events:

V_H4-31: GGG AGC ACC **TAC TAC** AAC CCG
g64: GGA Acc Aag --- **TAC** AAC CCG

V_H4-31: TCC ATC **AGC AGT** GGT GGT TAC
g187: TtC ATC **AGC** --- GGG GCT TAC

V_H4-59: TAC AGT GGG **AGC ACC** AAC TAC
g188: TAC Agg GGg --- **tCC** AAC TAC

V_H4-34: TAC TAC **TGG AGC TGG** ATC CGC
g80: agg TAC --- --- **TGG** ATC CGC

C Long duplication/insertion:

V_H4-31: CTGCACCTGCTCTCTGCTGCTCACTCA (CDR1).....
pg86: -----A-----T-----CCCTCTGCTGCTCACTCA-----
(80) (85)

Figure 4. The insertions and deletions are related to the surrounding DNA sequence. (A) The insertions involve repetitions of the immediately adjacent sequence. (B) The deletions are deletions of tandem repeats. (C) The 18-base insertion in clone pg86 is a duplication of the adjacent sequence. Nucleotides that mutated before the duplication/insertion are indicated. Sequence data available from EMBL/GenBank/DBJ under accession numbers AF013615 through AF013626.

the unmutated FM clones analyzed had insertions or deletions.

Other Insertions and Deletions into V_H Genes. We have observed similar instances of insertions and deletions into the coding regions of apparently functional immunoglobulin V genes, including: (a) a V_H6 IgM isolate containing a triplet duplication/insertion into the CDR1 in addition to several bp substitutions (Figs. 1 C and 4 A), which was derived from a human hybridoma secreting high affinity mAb against *Bordetella pertussis* (21, 22); (b) a 6-nucleotide insertion into the FW3 region of a mutated IgM V_H6 gene, representing the only insertion or deletion observed outside of the CDRs (Figs. 1 C and 4 A, clone tm121); and (c) an 18-nucleotide duplication/insertion into a human plasma cell cDNA transcript at the boundary between the FW1 and CDR1 (Figs. 1 C and 4 C), doubling the length of this hypervariable loop. The viability of clone pg86 was tested by expressing it in the baculovirus system in association with a κ light chain encoding construct (FS-6 κ ; Fig. 5). The efficient expression, secretion, and pairing with light chain in the baculovirus system suggest that the product of clone pg86 is a functional heavy chain despite the large duplication/insertion.

The Insertions and Deletions Are Related to the Surrounding Sequence. As shown in Fig. 4, the insertions reported are duplications of the immediately adjacent sequence, and the

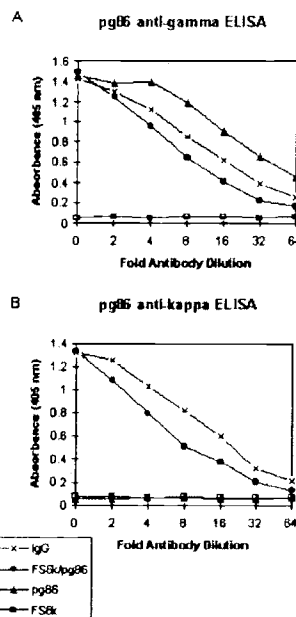


Figure 5. ELISA assays showing the expression of clone pg86 with a six amino acid insertion at the FW1/CDR1 junction. Clone pg86 (IgG heavy chain) was coexpressed with the κ light chain FS6 κ in insect cells using the baculovirus expression system. Expression of pg86 and its ability to pair with the FS6 κ light chain was tested using capture ELISAs. Wells were coated with goat anti-human IgG. Supernatants containing recombinant antibodies were added in serial twofold dilutions. Bound antibody was detected with phosphatase-conjugated goat anti-human IgG (A), and goat anti-human C κ (B).

deletions involve elements of repetitive tracts. In addition, a higher incidence of these events involve sequence motifs that resemble intrinsic hotspots of somatic hypermutation (12, 23-27): (a) four of eight events involved the serine codon AGC that has been reported as the "hottest" of hotspots (24-27) (Fig. 4, sequences HBp2, g187, g188, and g86); (b) two events involved TAC motifs (Fig. 4, g192 and g64); and (c) two events involved the motif AAC (Fig. 1, g144, and tm121). In general all of the clones found to contain insertions and deletions were highly mutated (Fig. 1). Several of these clones had bp substitutions clustered with the insertions or deletions (Figs. 1 and 4). The plasma cell transcript depicted in Fig. 4 C contained an 18-nucleotide insertion that duplicated the 5' adjacent sequence. The central nine nucleotides of the duplicated sequence form a partial palindrome (..GGtGaCtCC..). This clone was mutated (G to A at position 80 and an A to T at position 85) before the duplication/insertion event, as these mutations were perpetuated in the inserted sequence.

Discussion

Somatic modification of V genes encoding immunoglobulin and T cell receptors recapitulates most mechanisms observed in the evolutionary diversification of DNA: (a) V gene recombination, including imprecise junctions, P nucleotides, and untemplated N nucleotide addition; (b) gene conversion; and (c) bp substitutions in Ig somatic hypermutation. The insertion and deletion of nucleotides is another means for the evolutionary diversification of DNA, and has been proposed as an explanation for unusual V gene sequences in the past (Table 3). In this study, we show that insertions and deletions are associated with the somatic hypermutation process.

Complexities of the Analysis of Insertions and Deletions into V Genes. The formal characterization of these events has been a daunting task because of their low frequency, and the complexity of the germline V_H repertoire. According to our study, these events occur in <2% of somatically mutated clones. As shown in Fig. 2, the primary variability between V_H4 family members is 3–6-bp size variances in the CDR1s, which is comparable to the short insertions and deletions that we attribute to somatic hypermutation (in selected B cell populations). The similarity between evolutionary diversity and somatic diversification was expected, as the molecules are likely subject to the same functional and structural constraints. This has made it difficult to determine whether these events were generated somatically, versus germline encoded, or if they were artifacts of the experimental system: they could result from homologous recombination between alternate alleles or imperfect recombination between identical alleles, or they could have occurred during B cell replication independent of somatic hypermutation. In fact, V_H genes may exhibit particularly unstable sequence characteristics evolved to help support both germline diversity and the generation of somatic mutations, as suggested by the identification of intrinsic hotspots of somatic hypermutation within the CDRs of V genes (25, 26). Perhaps the area of greatest contention in this complex system remains the possibility that these low frequency events are artifacts of the experimental manipulations performed, the AMV-RT, Taq, or PFU polymerases, and/or the cloning in *E. coli*.

The Insertion/Deletion Events Are The Result of the Somatic Hypermutation Process. Our system addresses several key issues that associate the occurrence of insertions and deletions to the somatic hypermutation process. (a) Six of the nine insertions/deletions were identified within the V_H4 gene repertoire of a single tonsil, providing an experimental system that could be characterized extensively as described below. (b) All of the insertion/deletion events reported involved triplets or multiples of triplets, leaving the transcripts in frame and therefore functional, and eight of nine events reported were localized to the CDRs. As with somatic point mutations, no insertions or deletions were observed in the 80 to 120 nucleotides of constant region ($C\mu$ or $C\gamma$) DNA sequenced with each cDNA clone. These hallmarks of somatic hypermutation and selection argue strongly that these events are not artifacts. (c) The B cells analyzed were processed and separated into highly pure, mutated B cell populations including GC (IgD^-CD38^+) and memory (IgD^-CD38^-) B cells, and an unmutated FM B cell population (IgD^-CD38^-), making it possible to focus our analysis on the mutated populations and use the unmutated population as a negative control, which in turn allows the statistical association of the observed insertion and deletions to the somatic hypermutation process ($P = 0.014$). In addition, the isolation of four of the insertion/deletion events from memory B cells provides evidence that these events did not result from artifacts related to contamination from endonucleolytically cleaved DNA from the apoptotic GC cells. (d) Seven of nine events re-

ported in this study involved γ heavy chains that contain nearly twice the mutations of μ heavy chains (4), further correlating the events described here to somatic hypermutation. (e) As discussed below, the insertion/deletion events described tended to involve sequence motifs resembling previously described hotspots of somatic hypermutation, providing evidence that these events occur by the same process. (f) Finally, we extensively analyzed the V_H4 gene family of the tonsil donor at both the expressed and genomic levels, facilitating the assignment of the insertions/deletions as somatic rather than germline encoded. 6 of the clones with insertions and deletions were unique among 395 V_H4 cDNA clones sequenced from a single tonsil, including many independent isolates of each of the V_H4 genes expressed (Table 1). In addition, we were unable to isolate genomic templates for any of the insertion or deletion events either by PCR or through the extensive characterization of the genomic V_H4 repertoire of the tonsil donor (Table 1). Templating of these events from any other V_H gene family can also be ruled out as members of the seven human V_H gene families differ significantly in the CDR sequences where the events described had occurred.

Structural and Functional Considerations of Insertions and Deletions into V_H Genes. The events involving the insertion or deletion of a single amino acid from the CDR1 or CDR2 would not be expected to profoundly alter the backbone structure of these molecules, as the CDRs are the most malleable portions of antibodies. The clone g80 has two of the five amino acids that are customarily considered its CDR1 deleted, leaving only three amino acids to form this hypervariable loop (Fig. 1 B). Thus, this is one of the shortest CDR1s reported to date. The clone tm121 has two amino acids inserted into the FW3 region. The portion of the FW3 where this insertion occurred is believed to be solvent exposed and corresponds to the region where the B cell superantigen staphylococcal protein A binds to most V_H3 -encoded Ig molecules (28); therefore, it is likely that the insertion into this V_H6 clone can be tolerated as a loop or bulge on the molecule's surface. The most complex structural change observed in our study involved clone pg86, with a six amino acid insertion at the FW1/CDR1 junction that would presumably double the length of this hypervariable loop and require dramatic structural accommodation. However, we were able to express this heavy chain and found it paired with light chain, indicating that it is likely functional (Fig. 5). The clone HBp2, containing a triplet insert into its CDR1, is particularly interesting because it has a known specificity. This V_H6 gene was isolated from a human B cell hybridoma with anti-*Bordetella pertussis* specificity (21, 22). Clone HBp2 has also been expressed in the baculovirus system and is fully functional. We are currently performing mutational analysis of this heavy chain molecule to determine if the additional inserted amino acid plays a role in the affinity and/or specificity of this antibody.

Analysis of Insertions and Deletions Reported in the Literature. Various groups have reported a number of insertion and deletion events (Table 3). Virtually all of the insertions

Table 3. Insertions and Deletions into Somatically mutated V Genes Reported in the Literature

| Name | Source | Ins/Del (position) | Relation to surrounding sequence | References |
|--|--|--|---|------------|
| <i>Selected populations or coding regions:</i> | | | | |
| L4-le | Human V _H 4-34 (4.21) | ACC insert (within CDR2) | 4-34: AGC ACC AAC (RT) L4-le: AGC ACC ACC AAC (RT) | 38 |
| 3B62 | Murine V _H 186.2 | GTT deletion (CDR2) | VH186.2: AGT GGT GGT ACT (RT) 3B62: AGT GGT ACT | 39 |
| <i>Unselected populations or untranslated regions:</i> | | | | |
| 3B62 | Murine V _H 186.2/D/JH2 to JH4 | ACT deletion (3' untranslated) | GL: GTG ACT ACT TTG (RT) 3B62: GTG ACT TTG | 40 |
| 3B44 | Murine V _H 186.2 | 4 single-base deletions (leader intron) | VH186.2: GGC, GGT (RT) 3B62: G C, G T | 39, 40 |
| M167 | Murine VH107/DFL16.1/JH1 | 2 single-base insertions (leader intron) | Other 2 events unrelated GL: ATAG AAGATTAGTAG (RT) M167: ATAGTAAGATTAGTAG | 41 |
| | | (3' untranslated) | GL: TTC AGTCATGAAGGA (RT) M167: TTGAAGGTCATGAAGGA | 29, 41* |
| | | 5 single-base deletions | GL: GCTTTC TGTA...CCCAGAAAAGA M167: GCTTTC TGTA...CCCAG AAAAGA (IR) | 41 |
| | | (all in 3' untranslated region) | GL: CTTTTCTT (RT) M167: CTTTT TCTT | |
| | | | GL: AGATTTTAC (RT) M167: AGAccT AC | |
| | | | GL: TCATTGG (RT) M167: TCAT GG | |
| | | | GL: GTGACTACTTGACTACTG (RT) M167: GT ACTACTTGACTACTG | |

Continued

Table 3. Continued

| Name | Source | Ins/Del (position) | Relation to surrounding sequence | Reference |
|--|--|--|---|-----------|
| M603 | Murine V _H S10 7/DFL16.1/JH1 | TA deletion (3') GTGT deletion (leader intron) | None found (possible hotspot) GL: TCTGTGTGTGTAT (RT) M167: TCTGTGT GTAT GL: TTTTCCTTGTCTTGT (RT) M603: TTTTCCTTG CTTGTTTT GL: GCATTTCTAAATAAGTTGAGGA (IR) M603: GCATTTCTA AAGTTGAGGA GL: AAACGGGAATC (RT) MC101: AAAC GAATC GL: TTTGAAGATAAA (RT) m511: TTT GATAAA GL: AGGGACACCAAGTGTGTACAC (IL) H37: AGGGc GTACAC No good relationship to surrounding sequence CTTTGAAGAT... (N30)... CAGATCAAAG (Repeats form ends of deleted "loop") (IL) No relation, however, event followed the proposed hotspot motif TAC | 42, 41 |
| MC101 | Murine V _H Q52/D/J _H 3 | GG deletion (3' untranslated) | | 42, 43 |
| M511 | Murine V _K 167/J _K 5 | GAA deletion (3' untranslated) | | 42, 44 |
| H37-65 | Murine V _K V _K 21E/J _K 1-J _K 2 | 11 base deletion (J _K 1/J _K 2 intron) | | 45 |
| 296.4C11, 253.12D3 2G7 | Murine J _K C intron Murine transgene | 7 base deletion and a 154 base deletion single base deletion, and a 49 nucleotide deletion | | 46 47 |
| 85κ | Human myeloma V _K genes | single-base (T) insertions into the CDRI/FW2 junction rendering genes out-of-frame | | 48 |
| HF-1 ⁺ clone A6: several | Human lymphoma (J _H untranslated) | AG insertion into V _H 3' untranslated region 30 base deletion | Consensus [†] : GGGGCAG GGC (RT) clone A6: GGGGCAGAGGGC No association | 49 |

RT, repetitive tract; IR, inverted repeat (loop with local DNA); IL, internal loop. *Secondary structure reported by Golding et al. (29). [†]This study is difficult to interpret in the context of the current report as the genomic J_H locus was not available. The 10 clones were only 80% homologous to the closest J_H locus reported in the literature with most alterations being similar between all of the isolates. Therefore, only 2 of the 26 proposed insertions/deletions can be attributed to somatic mutation with certainty, as they were unique to the consensus of the individual clones.

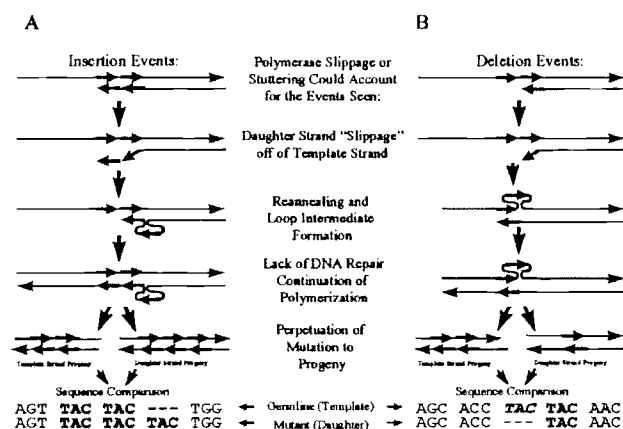


Figure 6. Proposed mechanism causing insertion/deletion events: polymerase slippage. This Figure is based on model a of Streisinger et al. (30) and Ripley (31). The same model can account for both (A) insertions and (B) deletions.

and deletions reported from somatically mutated V genes involved the untranslated regions or occurred in silent passenger transgenes. 19 out of 25 insertions or deletions into somatically mutated genes involved predominantly repetitive elements, or in several cases other sequence patterns associated with secondary structures such as internal homologies or inverted repeats (Table 3). With the inclusion of the 9 events described in this work, 28 out of 34 insertions and deletions involved such elements. Thus, the proximity of sequence elements that can be predicted to cause secondary structural changes in the DNA seems to be a hallmark of insertions and deletions into somatically mutated V_H genes.

A Model for the Occurrence of Insertions and Deletions during Somatic Hypermutation. The evidence for the involvement of DNA secondary structure in the production of insertion or deletion mutations during somatic hypermutation, as suggested in 1986 by Golding et al. (29), now seems unequivocal. The insertions and deletions described in our study, and those illustrated in Table 3, occur in a predictable fashion, involving sequence motifs that could form loop intermediates reminiscent of the replication slippage model of Streisinger et al. (30) and Ripley and Glickman (for review see 31) as presented in Fig. 6. Such mutations are postulated to occur when DNA polymerase slips or stutters and the newly synthesized strand shifts on the template and reanneals to an adjacent repetitive element, producing unpaired loop intermediates localized to one or the other strands. If this unpaired loop intermediate is not repaired then it will be perpetuated as an insertion of an instance of the repetitive element if in the daughter strand, or a deletion if in the template strand.

A Possible Correlation to Intrinsic Hotspots. A higher frequency of somatic hypermutation has been reported to occur at sequence motifs referred to as intrinsic hotspots (for review see reference 12). Interestingly, every insertion/deletion event reported in our study resembled one of these hotspots (AGC, TAC, and AAC; references 12 and 27; Fig.

4). The analysis of selected populations may have influenced this tendency because seven out of eight of these events occurred in the CDRs where it has been shown that hotspot motifs are preferentially found (25, 26). Furthermore, only a weak correlation to hotspots could be found for the previously reported insertions/deletions involving unselected regions of V loci (Table 3). However, the single event found in this analysis that occurred outside of the CDRs in FW3 (clone tm121, Figs. 1 C and 4 A), also involved a tandem of possible hotspots (AAG, AAC). A more extensive and directed analysis is required to fully address this issue.

Implications for the Molecular Mechanism of Somatic Hypermutation. The instability of repetitive tracts during DNA replication is a hallmark of defects in postreplicative mismatch repair (33), and the locus-specific downregulation of DNA mismatch repair in response to UV irradiation has recently been reported for immunoglobulin V_H genes in freshly sorted GC B cells (CD38⁺IgD⁻) compared to mantle zone B cells (CD38⁻IgD⁺; reference 34). In a recent study by Tran et al. (35), it was shown that tract instability of homonucleotide runs associated with mismatch repair defects occur more frequently in long than in short runs. These authors suggested that if loop intermediates occur in long repetitive tracts (>8 bp for a homonucleotide run) they could involve a distal repetitive element out of reach of the polymerase proofreading activity and only be subjected to mismatch repair. However, for short repetitive tracts, as for the events reported in this analysis, loop intermediates can only occur proximal to the polymerase complex and are therefore subjected to both polymerase proofreading and mismatch repair mechanisms.

All 9 events in this analysis, and 19 out of 25 events from the literature (28 out of 34 insertions and deletions reported), appeared to result from secondary structural intermediates. Loop intermediates proximal to the polymerase complex during DNA polymerization should be repaired by the polymerase proofreading mechanisms immediately, or by the postreplicative DNA repair systems. This analysis suggests the following characteristics for the polymerization process during somatic hypermutation. (a) The polymerase interacts with the V locus in a particularly unstable or "loose" fashion, especially when hotspot motifs or elements capable of forming secondary structures are encountered, allowing bp substitutions in most instances, and insertions or deletions via polymerase slippage at a much lower frequency; (b) it has limited proofreading capabilities; and (c) there is a downregulation of postreplicative mismatch repair. An efficient means to downregulate mismatch repair during somatic hypermutation could be through the lack of differentiation of the template and progeny strands for the mismatch repair system; lack of strand differentiation has been shown to increase the rate of mutations introduced (36). Such a system would be advantageous for the locus-specific V gene somatic hypermutation in that it could involve alterations of a single enzymatic complex (polymerase complex) rather than multiple systems (proofreading and mismatch repair). Another system, which would have the

same advantage, i.e., the alteration of a single complex, would be the alteration of a DNA repair system such as transcription-coupled repair to be the somatic mutator, as suggested in recent studies (13). Alternatively, the insertions and deletions might result solely from a downregulation of postreplicative mismatch repair at the V locus in the rapidly proliferating centroblasts that are undergoing somatic hypermutation or due to a polymerase enzyme with such a high fault rate as to overwhelm any repair.

All currently accepted models of somatic hypermutation, whether related to DNA excision-repair-like systems or transcription-repair, or to DNA polymerization or reverse transcription, involve transcriptional activation involving *cis*-factors in the V locus (enhancers, etc.) followed by the activity of unknown polymerase enzymes of some type. This analysis does not refute or corroborate any of these models directly, but it does provide further characterization of the polymerization system involved, based on the types

of mutations observed and on the molecular biology that is known to cause such mutations. This analysis and the model presented here provide further information or criteria to be contemplated as the various possible polymerase systems involved are considered.

Conclusions. Insertions and deletions into immunoglobulin V_H genes during somatic hypermutation are additional means by which the immunoglobulin repertoire can be diversified. These events display characteristics supporting models of somatic hypermutation involving a particularly unstable or error-prone polymerase to allow the introduction of mutations, and involving the downregulation of DNA repair to allow the perpetuation of these mutations. Additionally, we show that these events tend to involve sequence motifs resembling intrinsic hotspots of somatic hypermutation, suggesting that the polymerase complex is destabilized in a sequence-specific manner to allow preferential mutation at these sequence elements.

We are grateful to Yucheng Li, Fang Zhao, Steve Scholl, Carol Williams, Shirley Hall, and Robin Wray, all of whom provided unprecedented technical assistance for various aspects of this work. We also thank Kimble Frazer for excellent discussions.

These studies were supported by a grant from the National Institutes of Health (AI-12127). J. Donald Capra holds the Edwin L. Cox Distinguished Chair in Immunology and Genetics.

Address correspondence to Dr. Virginia Pascual, Molecular Immunology Center, Department of Microbiology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, Texas 75235-9140. Phone: 214-648-1918; Fax: 214-648-1915; E-mail: vpascu@mednet.swmed.edu

Received for publication 14 July 1997 and in revised form 1 October 1997.

References

- Clark, E.A., and J.A. Ledbetter. 1994. How B and T cells talk to each other. *Nature*. 367:425-428.
- Dorken, B., P. Moller, A. Pezzutto, R. Schwartz-Albiez, and G. Moldenhauer. 1989. B-cell antigens. In *Leukocyte Typing IV*. A.J. McMichael, editor. Oxford University Press, London. pp. 131-140.
- Ling, N.R., I.C.M. MacLennan, and D.Y. Mason. 1987. B-cell and plasma cell antigens: new and previously defined clusters. In *Leukocyte Typing III*. A.J. McMichael, editor. Oxford University Press, London. pp. 302-308.
- Pascual, V., Y.J. Liu, A. Magalski, O. de Bouteiller, J. Banchereau, and J.D. Capra. 1994. Analysis of somatic mutation in five B cell subsets of human tonsil. *J. Exp. Med.* 180: 329-339.
- Liu, Y.J., O. de Bouteiller, C. Arpin, F. Briere, L. Galibert, S. Ho, H. Martinez-Valdez, J. Banchereau, and S. Lebecque. 1996. Normal human IgD⁺IgM⁺ germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity*. 4:603-613.
- Denepoux, S., D. Razanajaona, D. Blanchard, G. Meffre, J.D. Capra, J. Banchereau, and S. Lebecque. 1997. Induction of somatic mutation in a human B cell line in vitro. *Immunity*. 6:35-46.
- Berek, C. 1993. Somatic mutation and memory. *Curr. Opin. Immunol.* 5:218-222.
- Kelsoe, G. 1996. Life and death in germinal centers (redux). *Immunity*. 4:107-111.
- Liu, Y.J., G.D. Johnson, J. Gordon, and I.C.M. MacLennan. 1992. Germinal centres in T-cell-dependent antibody responses. *Immunol. Today*. 13:17-21.
- MacLennan, I.C.M., Y.J. Liu, and G.D. Johnson. 1992. Maturation and dispersal of B-cell clones during T cell-dependent antibody responses. *Immunol. Rev.* 126:143-161.
- McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA*. 81:3180-3186.
- Neuberger, M.S., and C. Milstein. 1995. Somatic hypermutation. *Curr. Opin. Immunol.* 7:248-254.
- Storb, U. 1996. The molecular basis of somatic hypermutation of immunoglobulin genes. *Curr. Opin. Immunol.* 8:206-214.
- Liu, Y.J., O. de Bouteiller, C. Arpin, I. Durand, and J. Banchereau. 1994. Five human mature B cell subsets. *Adv. Exp. Med. Biol.* 355:289-296.
- Marks, J.D., M. Tristem, A. Karpas, and G. Winter. 1991. Oligonucleotide primers for polymerase chain reaction amplification of human immunoglobulin variable genes and design of family-specific oligonucleotide probes. *Eur. J. Immunol.* 21:985-991.
- Sanz, I., P. Kelly, C. Williams, S. Scholl, P. Tucker, and J.D. Capra. 1989. The smaller human V_H gene families display remarkably little polymorphism. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:3741-3448.
- Potter, K.N., Y.C. Li, and J.D. Capra. 1994. The cross-reac-

- tive idiotopes recognized by the monoclonal antibodies 9G4 and LC1 are located in framework region I of two non-overlapping subsets of human V_H4 family encoded antibodies. *Scand. J. Immunol.* 40:43-49.
18. Tomlinson, I.M., G. Walter, J.D. Marks, M.B. Llewelyn, and G. Winter. 1992. The repertoire of human germline V_H sequences reveals about fifty groups of V_H segments with different hypervariable loops. *J. Mol. Biol.* 227:776-798.
 19. Umar, A., and P.J. Gearhart. 1995. Reciprocal homologous recombination in or near antibody VDJ genes. *Eur. J. Immunol.* 25:2392-2400.
 20. Ford, J.E., M.G. McHeyzer-Williams, and M.R. Lieber. 1994. Chimeric molecules created by gene amplification interfere with the analysis of somatic hypermutation of murine immunoglobulin genes. *Gene*. 142:279-283.
 21. Andris, J.S., B.R. Brodeur, and J.D. Capra. 1993. Molecular characterization of human antibodies to bacterial antigens: utilization of the less frequently expressed V_H2 and V_H6 heavy chain variable region gene families. *Mol. Immunol.* 30:1601-1616.
 22. Brodeur, B.R., J. Hamel, D. Martin, and P. Rondeau. 1991. Biological activity of a human monoclonal antibody to *Bordetella pertussis* lipooligosaccharide. *Hum. Antib. Hybrid.* 2:194-199.
 23. Betz, A.G., C. Rada, R. Pannell, C. Milstein, and M.S. Neuberger. 1993. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc. Natl. Acad. Sci. USA*. 90:2385-2388.
 24. Rogozin, I.B., and N.A. Kolchanov. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta*. 1171:11-18.
 25. Goyenchea, B., and C. Milstein. 1996. Modifying the sequence of an immunoglobulin V-gene alters the resulting pattern of hypermutation. *Proc. Natl. Acad. Sci. USA*. 93:13979-13984.
 26. Wagner, S.D., C. Milstein, and M.S. Neuberger. 1995. Codon bias targets mutation. *Nature*. 376:732-733.
 27. Smith, D.S., G. Creadon, P.K. Jena, J.P. Portanova, B.L. Kotzin, and L.J. Wysocki. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol.* 156:2642-2652.
 28. Potter, K.N., Y. Li, and J.D. Capra. 1996. Staphylococcal protein A simultaneously interacts with framework region 1, complementarity-determining region 2, and framework region 3 on human V_H3-encoded Igs. *J. Immunol.* 157:2982-2988.
 29. Golding, G.B., P.J. Gearhart, and B.W. Glickman. 1987. Patterns of somatic mutations in immunoglobulin variable genes. *Genetics*. 115:169-176.
 30. Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harbor. Symp. Quant. Biol.* 31:77-84.
 31. Ripley, L.S. 1990. Frameshift mutation: determinants of specificity. *Annu. Rev. Genet.* 24:189-213.
 32. Modrich, P., and R. Lahue. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65:101-133.
 33. Strand, M., T.A. Prolla, R.M. Liskay, and T.D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*. 365:274-276.
 34. Fairhurst, R.M., Y. Valles-Ayoub, M. Neshat, and J. Braun. 1996. A DNA repair abnormality specific for rearranged immunoglobulin variable genes in germinal center B cells. *Mol. Immunol.* 33:231-244.
 35. Tran, H.T., J.D. Keen, M. Kricker, M.A. Resnick, and D.A. Gordenin. 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell Biol.* 17:2859-2865.
 36. MacPhee, D.G. 1996. Mismatch repair as a source of mutations in non-dividing cells. *Genetica (The Hague)*. 97:183-195.
 37. Matsuda, F., and T. Honjo. 1996. Organization of the human immunoglobulin heavy-chain locus. *Adv. Immunol.* 62:1-29.
 38. Dunn-Walters, D.K., P.G. Isaacson, and J. Spencer. 1997. Sequence analysis of human IgV_H genes indicates that ileal lamina propria plasma cells are derived from Peyer's patches. *Eur. J. Immunol.* 27:463-467.
 39. Both, G.W., L. Taylor, J.W. Pollard, and E.J. Steele. 1990. Distribution of mutations around rearranged heavy-chain antibody variable-region genes. *Mol. Cell Biol.* 10:5187-5196.
 40. Allen, D., T. Simon, F. Sablitzky, K. Rajewsky, and A. Cumano. 1988. Antibody engineering for the analysis of affinity maturation of an anti-hapten. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1995-2001.
 41. Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. *Cell*. 27:573-581.
 42. Lebecque, S.G., and P.J. Gearhart. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med.* 172:1717-1727.
 43. Kataoka, T., T. Nikaido, T. Miyata, K. Moriwaki, and T. Honjo. 1982. The nucleotide sequences of rearranged and germline immunoglobulin V_H genes of a mouse myeloma MC101 and evolution of V_H genes in mouse. *J. Biol. Chem.* 257:277-285.
 44. Gearhart, P.J., and D.F. Bogenhagen. 1983. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc. Natl. Acad. Sci. USA*. 80:3439-3443.
 45. Rickert, R. and S. Clarke. 1993. Low frequencies of somatic mutation in two expressed V kappa genes: unequal distribution of mutation in 5' and 3' flanking regions. *Int. Immunol.* 5:255-263.
 46. Weber, J.S., J. Berry, S. Litwin, and J.L. Claflin. 1991. Somatic hypermutation of the J_C intron is markedly reduced in unrearranged kappa and H alleles and is unevenly distributed in rearranged alleles. *J. Immunol.* 146:3218-3226.
 47. Rogerson, B., J. Hackett, Jr., A. Peters, D. Haasch, and U. Storb. 1991. Mutation pattern of immunoglobulin transgenes is compatible with a model of somatic hypermutation in which targeting of the mutator is linked to the direction of DNA replication. *EMBO (Eur. Mol. Biol. Organ.) J.* 10:4331-4341.
 48. Kosmas, C., N.A. Viniou, K. Stamatopoulos, N.S. Courtenay-Luck, T. Papadaki, P. Kollia, G. Paterakis, D. Anagnostou, X. Yataganas, and D. Loukopoulou. 1996. Analysis of the kappa light chain variable region in multiple myeloma. *Br. J. Haematol.* 94:306-317.
 49. Wu, H.Y., and M. Kaartinen. 1995. The somatic hypermutation activity of a follicular lymphoma links to large insertions and deletions of immunoglobulin genes. *Scand. J. Immunol.* 42:52-59.

Functional Consequences of Insertions and Deletions in the Complementarity-determining Regions of Human Antibodies*

Received for publication, August 16, 2002, and in revised form, September 16, 2002
Published, JBC Papers in Press, September 16, 2002, DOI 10.1074/jbc.M208401200

Johan Lantto and Mats Ohlin†

From the Department of Immunotechnology, Lund University, P.O. Box 7031, S-220 07 Lund, Sweden

Insertions and deletions of nucleotides in the genes encoding the variable domains of antibodies are natural components of the hypermutation process, which may expand the available repertoire of hypervariable loop lengths and conformations. Although insertion of amino acids has also been utilized in antibody engineering, little is known about the functional consequences of such modifications. To investigate this further, we have introduced single-codon insertions and deletions as well as more complex modifications in the complementarity-determining regions of human antibody fragments with different specificities. Our results demonstrate that single amino acid insertions and deletions are generally well tolerated and permit production of stably folded proteins, often with retained antigen recognition, despite the fact that the thus modified loops carry amino acids that are disallowed at key residue positions in canonical loops of the corresponding length or are of a length not associated with a known canonical structure. We have thus shown that single-codon insertions and deletions can efficiently be utilized to expand structure and sequence space of the antigen-binding site beyond what is encoded by the germline gene repertoire.

Antibodies are highly specific receptors of the immune system that also have a great potential as reagents in biological chemistry and as therapeutic agents. The part of the antibody that makes contact with the antigen is comprised of two variable (V)¹ domains, the heavy (H) and the light (L), which both are made up of a two- β -sheet framework. From this framework, six complementarity-determining region (CDR) loops, three from the light domain and three from the heavy domain, protrude and make up the antigen-binding site (1, 2). Five of these CDR loops generally adopt only a limited number of backbone conformations, so-called canonical structures (reviewed in Ref. 3), which are determined by the lengths of the loops and by the presence of specific key residues. The antigen specificity of the binding site is mainly determined by the sequence and conformation of these CDR loops.

Antibody diversity is generated by the imprecise recombination of two or three sets of germline gene segments and by the

combination of different heavy and light domains (4). The diversity is further increased by the process of somatic hypermutation (5) and by receptor editing and revision (6). As the germline variable gene repertoire encodes a rather limited number of CDR loop lengths (IMGT, the international ImMunoGeneTics data base, Ref. 7), the number of observed canonical structures is similarly limited. However, it was recently discovered that B cells evolve the genes encoding immunoglobulin V domains not only by nucleotide substitution but also through an additional mechanism of insertion and deletion of nucleotides during the hypermutation process (8–11). This mechanism has the potential to expand the available repertoire of loop lengths and conformations if the insertions and deletions involve entire codons and occur at positions in the sequence that can tolerate such modifications. A number of examples of seemingly functional insertions and deletions in the CDR of both the heavy and light domains of human antibodies have in fact been encountered lately (Refs. 8 and 12 and references therein). Furthermore, we have recently discovered that human IGHV² germline genes carry features in CDR1 and CDR2 that make these regions particularly prone to deletions of entire codons (12).

The occurrence of insertions and deletions in antibody V genes is not only of fundamental interest but is also of biotechnological importance. It has been known for some time that the topography of the antigen-binding site is related to the size of the antigen (13–15). Three different types of binding sites have been described: cavity, groove, and planar, which roughly correspond to hapten, peptide, and protein, respectively. This relationship has been further investigated by Vargas-Madrado *et al.* (16), who have described a correlation between the length of the CDR loops and the antigen recognized. According to these findings, cleft-like binding sites that recognize small molecules are created by long loops (especially the CDRH2 and L1 loops), whereas planar-binding sites that are specific for large molecules are formed by short loops. In other words, by modifying the loop lengths of an antibody-binding site, it may thus be possible to design antibodies optimally suited for recognition of a particular class of antigen. Lamminmäki *et al.* (17) have in fact used this approach to modify a murine antibody specific for 17 β -estradiol. They introduced additional residues into CDR2 of the heavy domain and were able to improve the recognition of the antigen. This improvement was suggested to be the result of a deeper binding site, created through the extension of CDRH2, which better accommodated the hapten (17).

Despite the establishment of insertions and deletions as naturally occurring modifications of antibody sequences and the use of amino acid insertions for antibody engineering, little is still known about the functional consequences of such mod-

* This study was supported by BioInvent Therapeutic AB, the Swedish Research Council, and the Crafoord Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Tel.: 46-46-222-4322; Fax: 46-46-222-4200; E-mail: mats.ohlin@immun.lth.se.

¹ The abbreviations used are: V, variable; L, light; H, heavy; CDR, complementarity-determining region; scFv, single chain variable region fragment; FITC, fluorescein isothiocyanate; BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay; DSC, differential scanning calorimetry; PBS, phosphate-buffered saline.

² The immunoglobulin gene names used in this report are according to the official IMGT/HUGO nomenclature (IMGT, the international ImMunoGeneTics database, Ref. 7).

ifications. We have therefore created single-codon insertions and deletions as well as more complex modifications in the CDR of two human antibody single chain V region fragments (scFv) specific for a peptide and a hapten, respectively, and investigated the effects on antigen recognition, thermal stability, and protein folding. Our results demonstrate that single amino acid insertions in both CDRH1 and H2 and deletions in CDRH2 are usually well tolerated and permit production of folded proteins despite the fact that the modified loops carry amino acids that are disallowed at key residue positions in canonical loops of the corresponding length or do not take on a characteristic length of a known canonical structure. Modifications of this kind are in other words an efficient mode of expanding antibody sequence and structure space beyond what is encoded by the germline gene repertoire, which may enable targeting of novel or otherwise poorly immunogenic antigens.

EXPERIMENTAL PROCEDURES

Antibody Frameworks—The frameworks encoding the anti-cytomegalovirus scFv AE11F and the anti-fluorescein isothiocyanate (FITC) scFv FITC8 have been described elsewhere (18–20). The cloning and production of the AE11F and AE11F/3-20L1 scFv in *Pichia pastoris* has also been described (21).

Creation of Insertion and Deletion Variants—Mini-libraries of scFv genes carrying codon insertions at various positions were created by the use of overlap extension PCR with degenerate primers that introduced NNK codons. Variants with a deletion were similarly created with primers lacking one codon. The AE11F-based variants carrying CDRH1 sequences derived from the IGHV4 subgroup were created using the CDR-shuffling technique (22) essentially as described previously (21, 23).

Production and Purification of scFv Variants—The FITC8 scFv and all variant scFv were cloned into the pPICZα vector (Invitrogen) with C-terminal FLAG sequences (24) and produced in *P. pastoris* as described previously (21). The mini-libraries encoding AE11F and FITC8 variants were screened for scFv production or antigen binding according to the colony lift assay by McGrew *et al.* (25). Briefly, transformed *P. pastoris* colonies were lifted onto cellulose acetate filters (Pall Gelman Sciences, Ann Arbor, MI) and were grown on top of nitrocellulose filters, which were placed on methanol-containing plates. After 48 h of induction, scFv bound to the nitrocellulose filters were detected by a combination of anti-FLAG M2 antibody (Sigma) and rabbit anti-mouse Ig/horseradish peroxidase conjugate (DAKO A/S, Glostrup, Denmark) or FITC-biotin (Sigma) and streptavidin/horseradish peroxidase conjugate (DAKO A/S) using the ECL Plus™ Western blotting detection reagents (Amersham Biosciences) according to the manufacturer's recommendations. Single colonies were also picked and grown in liquid cultures to enable further characterization of the antigen binding properties (see below). In addition, a number of scFv variants were produced at a larger scale and purified as monomers. The AE11F-based variants were purified essentially as described previously (21), whereas the FITC8-based variants were purified by affinity chromatography on a Sepharose resin with FITC-conjugated bovine serum albumin (BSA) (kindly provided by Dr. B. Jansson, BioInvent Therapeutic AB, Lund, Sweden) followed by gel filtration as before.

Analysis of Antigen Recognition—The reactivity of the scFv variants with different antigens, both as crude expression supernatants and as purified monomers, was analyzed by enzyme-linked immunosorbent assay (ELISA) and by using the BiAcCore technology (BiAcCore AB, Uppsala, Sweden). The AE11F-based clones were tested on BSA, ovalbumin, streptavidin, and a biotinylated peptide that mimics the viral epitope (21) bound via streptavidin and the FITC8-based clones on BSA, streptavidin, FITC-BSA, FITC-biotin (bound via streptavidin), and a number of irrelevant BSA-coupled haptens obtained from Sigma or Biosearch Technologies Inc. (Novato, CA). The ELISA was performed according to standard protocols with anti-FLAG M2 (Sigma) and rabbit anti-mouse immunoglobulin/horseradish peroxidase conjugate (DAKO) to detect bound scFv. The BiAcCore measurements and the calculation of the reaction rate kinetics were performed essentially as described previously (21).

Differential Scanning Calorimetry (DSC)—DSC measurements were performed using a VP-DSC from Microcal Inc. (Northampton, MA) in the temperature range 20–90 °C at a heating rate of 60 °C/h. All measurements were performed in phosphate-buffered saline (PBS), pH 7.4, containing 0.02% sodium azide at protein concentrations between 0.1

and 0.2 mg/ml with PBS in the reference cell. Prior to protein *versus* PBS measurements, PBS *versus* PBS scans were performed.

CD Spectroscopy—CD spectra were recorded on a J-720 spectropolarimeter (Jasco Inc., Easton, MD) in a 2-mm cuvette at a protein concentration of 0.1 mg/ml in 50 mM sodium phosphate, pH 7.4. Each sample was scanned two to eight times from 250 to 200 nm at a scan speed of 10 nm/min, a resolution of 1 nm, a bandwidth of 1 nm, and a sensitivity of 20 millidegrees, and the scans were combined to produce the final spectrum. Data are presented as mean residue molar ellipticity, which was calculated using the mean residue weight of each scFv.

Sequencing and Canonical Structure Classification—The nucleotide sequences of the variant scFv clones were determined by automated DNA sequencing as described elsewhere (26) after isolation of the templates by direct PCR on *P. pastoris* colonies using vector-specific primers. In the case of the CDRH1-grafted clones, the origin of the CDR was determined using the IMGT/V-QUEST alignment tool at IMGT, the international ImMunoGeneTics data base (imgt.cines.fr and Ref. 7). All sequences were defined and numbered in accordance with the IMGT nomenclature and unique numbering (7). Complete sequences of the variant scFv from this study can be found in GenBank™ under accession codes AF543317–AF543349. The canonical structure classification was performed using the software implemented on the Antibodies – Structure and Sequence server (www.bioinf.org.uk/abs/chothia.html and Ref. 27).

RESULTS

The scFv Frameworks—The parent antibody frameworks used in this study are both of human origin although there are differences in the way they were obtained. The AE11F scFv was derived from a monoclonal antibody isolated from a cytomegalovirus-seropositive blood donor (18, 19). It originates from the IGHV3-30 and IGKV3-11 genes, which both have acquired a number of mutations (21). This scFv recognizes both intact glycoprotein B from cytomegalovirus and peptides mimicking the AD-2 epitope (21, 28). The hapten (FITC)-specific scFv FITC8 was derived from a synthetic scFv library, which had been constructed by shuffling of human CDR sequences into a single framework consisting of the human IGHV3-23 and IGLV1-47 genes (20). The CDR sequences utilized by this scFv originate from IGHV3-7 and IGHV3-23 in the case of CDRH1 and CDRH2, IGLV1-40 and IGLV1-40 or IGLV1-50 in the case of CDRL1 and CDRL2, and IGLV1-47 in the case of CDRL3. Except for the CDRL1 loop, which is one residue longer than the IGLV1-47 germline length, the CDR loops of the FITC8 scFv are of the same length as the loops normally encoded by the framework genes. As the structures of the two scFv have not been determined, the loop structures are unknown. However, by analyzing the deduced amino acid sequences using the tools at the Antibodies – Structure and Sequence server (27), the most similar of the observed canonical classes were identified (Table I).

Single-codon Insertions and Deletions—To determine the capability of the two antibody frameworks to tolerate length modifications in the CDR loops, we made single-codon insertions in CDRH1 and CDRH2 and a single-codon deletion in CDRH2. The modifications involved insertions after positions 31–33 in CDRH1, insertions after positions 57 and 58 in CDRH2, and a deletion at position 58 in CDRH2 (Fig. 1). All modifications were introduced at positions corresponding to the apices of the loops, *i.e.* the positions where the natural length variation occurs (31). A study of the IGHV germline gene repertoire has shown that these parts of the CDR carry repetitive sequence tracts, which naturally target them with deletions (and possibly also insertions) during the hypermutational process (12). Residues in these regions have also been shown to frequently make contact with the antigen in known antibody-antigen complexes (15), suggesting that modifications at the above mentioned positions will result in an expansion of structure space that is relevant for antigen recognition.

Libraries of scFv clones producing different insertion vari-

TABLE I

Examples of various single-codon modifications in scFv clones based on the AE11F and FITC8 frameworks, the canonical class belonging of the CDR loops, reactivity of the scFv with the original antigens as determined by ELISA or by BIAcore measurements, and the unfolding temperature of selected clones as determined by DSC

Modification refers to the nature of the changes in loop length; Ins indicates insertion, and Del indicates deletion. Numbering is according to the IMGT unique numbering (7). Canonical class indicates the combination of canonical structures of CDRH1, H2, and L1 as determined by automatic canonical structure classification (27). The altered canonical structure is indicated in *bold*. Antigen recognition: -, negative; ±, weakly positive; +, positive; ++, strongly positive.

| scFv clone | Modification | Canonical class | Antigen recognition | Unfolding temperature |
|--------------|------------------------|---------------------|---------------------|-----------------------|
| | | <i>H1-H2-L1</i> | | °C |
| AE11F | Original sequence | 1-3-2 | ++ | 62.4 |
| ASV18 | Ins Pro-31A | 2-3-2 | + | |
| ASV19 | Ins Asn-31A | 2-3-2 | + | |
| ASV43 | Ins Arg-31A | 2-3-2 | + | |
| ASV15 | Ins His-32A | 2-3-2 | + | |
| ASV37 | Ins Ile-32A | 2-3-2 | ± | |
| ASV39 | Ins Phe-32A | 2-3-2 | + | |
| ASV02 | Ins Phe-33A | 2-3-2 | ± | |
| ASV35 | Ins Asn-33A | 2-3-2 | - | |
| ASV07 | Ins Lys-57A | 1-U ^a -2 | ++ | 63.0 |
| ASV08 | Ins Ile-57A | 1-U-2 | + | |
| ASV28 | Ins Thr-57A | 1-U-2 | ++ | |
| ASV05 | Ins Glu-58A | 1-U-2 | ++ | |
| ASV10 | Del Val-58 | 1-1-2 | ± | 61.9 |
| AE11F/3-20L1 | Ins CDRL1 ^b | 1-3-6 ^c | + | 62.3 |
| | | <i>H1-H2</i> | | |
| FITC8 | Original sequence | 1-3 | ++ | 63.4 |
| FSV71 | Ins Ser-31A | 2-3 | + | |
| FSV73 | Ins His-31A | 2-3 | ± | |
| FSV76 | Ins Arg-31A | 2-3 | + | |
| FSV81 | Ins Asn-32A | 2-3 | + | |
| FSV84 | Ins Pro-32A | 2-3 | ++ | 61.7 |
| FSV85 | Ins Arg-32A | 2-3 | + | |
| FSV91 | Ins Leu-33A | 2-3 | + | |
| FSV93 | Ins His-33A | 2-3 | + | |
| FSV96 | Ins Tyr-33A | 2-3 | + | |
| FSV51 | Ins Ser-57A | 1-U | ++ | |
| FSV52 | Ins Ala-57A | 1-U | ++ | |
| FSV56 | Ins Leu-57A | 1-U | ++ | |
| FSV43 | Ins Thr-58A | 1-U | ++ | 61.1 |
| FSV46 | Ins Arg-58A | 1-U | ± | |
| FSV61 | Del Gly-58 | 1-1 | + | 63.9 |

^a U indicates that the canonical structure of the created loop length is currently unknown.

^b See text for details regarding the insertion.

^c The automatic canonical class algorithms failed to unambiguously predict a structure for the CDRL1 loop of this scFv. Similarities in length and sequence with Fab 1f7 (PDB entry 1fig) suggest that the loop belongs to canonical structure class 6 (30).

ants were screened directly by the use of a colony lift assay (25). This analysis showed that ~95% of the clones based on the FITC8 framework had retained their specificity for FITC (data not shown). The libraries based on the AE11F framework were screened for the production of FLAG-carrying proteins, and a similar ratio of clones positive for scFv production was obtained (data not shown). Both positive and negative clones from each library were sequenced to determine the nature of the modifications, and the analysis showed that a wide range of amino acids was inserted at the intended positions. To determine the effect of these length modifications on the structure of the targeted loops, the most similar canonical structures were identified by the automatic canonical structure classification (27). A number of examples from each insertion library and the deletion variants are presented in Table I.

As the AE11F-based libraries were only tested for the production of FLAG-tagged proteins, they had to be characterized further to determine whether the scFv were functionally folded. This was done by analyzing the antigen-binding properties of the modified clones. Although changes in loop structure may be associated with a loss of antigen recognition, specific recognition of an antigen will confirm that the polypeptide chain is correctly folded as this is a requirement for it to function as a framework for the antigen-binding site. Analysis of expression supernatants of randomly picked clones (including the deletion variants) by ELISA or by using the BIAcore

technology confirmed the above finding that the majority of the FITC8-based clones recognized the original antigen. Importantly, this analysis showed that most of the AE11F-based clones had also retained their specificity for the original viral antigen (Table I). Furthermore, when tested for binding to a number of irrelevant antigens (see "Experimental Procedures"), none of the clones displayed any cross-reactivity (data not shown), demonstrating that the modified scFv clones retained a high degree of specificity for the original antigens and therefore likely also assumed a correct immunoglobulin fold.

A number of clones of each specificity, chosen to exemplify the different modifications, were produced at a large scale to study the interaction with the original antigens in detail and determine the stability of the purified proteins. BIAcore measurements with the purified monomers of the ASV07, ASV10, ASV35, FSV43, FSV61, and FSV84 clones confirmed the previously obtained results with crude expression supernatants (Table I and Fig. 2). Furthermore, evaluation of the reaction rate kinetics with the original antigen showed that the modifications did not affect the dissociation rates of the FITC8-based clones to any greater extent (Fig. 2B). The thermal stability of the purified monomers was determined by DSC, and all tested clones displayed unfolding temperatures very similar to the parent scFv (Table I), further verifying that the IGHV3-derived antibody frameworks tolerate single-codon insertions and deletions in CDRH1 and H2 very well.

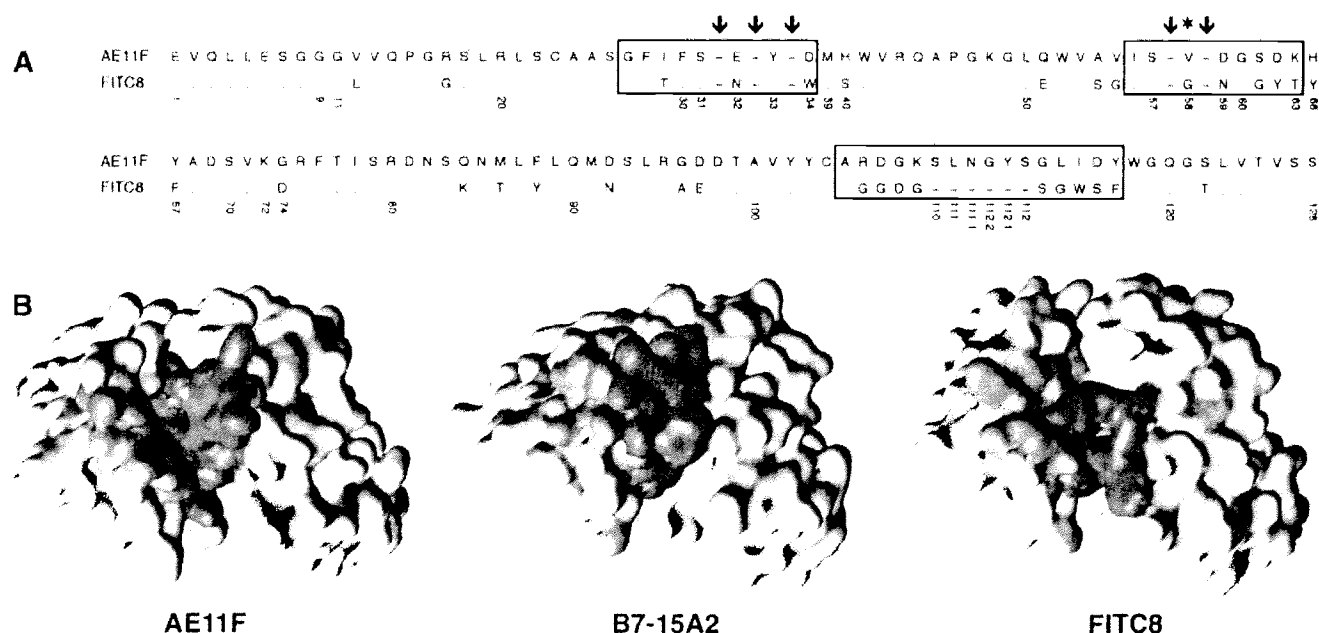


FIG. 1. Sequences and structures of the scFv frameworks used for production of insertion and deletion variants. A, alignment of the deduced amino acid sequences of the heavy V domains of the AE11F and FITC8 scFv. CDR-IMGT are boxed, and the location of the insertions and the deletion made in this study are indicated by arrows and an asterisk, respectively. Amino acid numbering according to the IMGT unique numbering is shown below the sequences. B, location of the affected sequences as indicated on a structure model of AE11F, which was generated using the WAM algorithm (29), a determined structure of the protein-specific antibody B7-15A2 (Protein Data Bank entry 1a9k), which originates from a highly related IGHV gene and has a CDRH3 of the same length as AE11F, and a structure model of FITC8 (20). CDRH3 is shown in red, whereas residues immediately adjacent to the single-codon insertions and the deletion made in this study are highlighted in blue (residues 31–34 in CDRH1) and green (residues 57–59 in CDRH2), respectively.

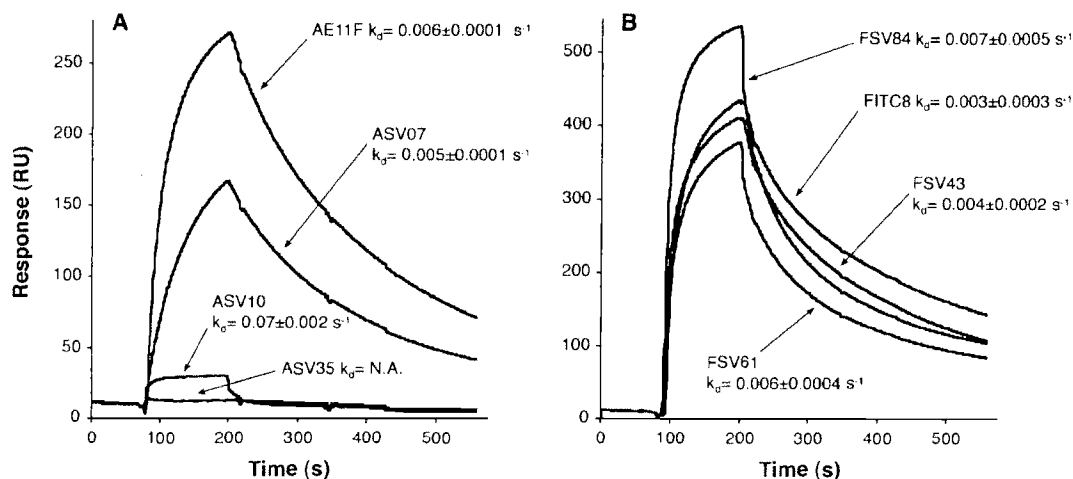


FIG. 2. The single-codon modifications did not affect the reaction rate kinetics of the FITC8-based scFv variants to any greater extent. Representative BIAcore sensorgrams of AE11F-based scFv and FITC8-based scFv analyzed on streptavidine-bound viral peptide (A) or streptavidine-bound FITC (B), respectively. Dissociation rate constants (k_d) were calculated from multiple measurements and are presented as the mean value \pm S.E. N.A., not applicable; RU, resonance units.

As insertions and deletions have been demonstrated to occur naturally in both heavy and light domain V genes (8), we decided to extend this study and also evaluate the stability of a previously produced AE11F-based scFv variant with an insertion in CDRL1 (AE11F/3-20L1) (21). The modified CDRL1 of this scFv is identical, except for an additional serine residue, to the germline gene from which AE11F originates. This clone has also been demonstrated to recognize both the epitope-mimicking peptide and intact, recombinant glycoprotein B, albeit with a lower affinity than the affinity matured AE11F scFv (21, 32). The thermal stability of the AE11F/3-20L1 scFv was determined as before after purification of monomeric scFv, and the unfolding temperature was found to be similar to that of the

original scFv (Table I), thus indicating that not only heavy but also light domain CDR tolerate modifications of this nature well.

Grafting of CDRH1 Loops from Distantly Related IGHV Genes—As all of the insertions and deletions described so far were introduced at the tips of the hypervariable loops, the parts of the immunoglobulin fold that best can be expected to accommodate such modifications, we decided to introduce more extensive modifications to investigate the effect of such changes of antibody sequence and structure. These modifications were introduced into and immediately adjacent to CDRH1 of the AE11F framework by the CDR-shuffling technique (22) using CDR sequences isolated from activated human B cells. Sequences originating from the IGHV4 subgroup were chosen for

TABLE II

Deduced amino acid sequences, germline gene origin, and canonical structure class belonging of the CDRH1 loops of the AE11F scFv and the CDRH1-grafted variants of this

Amino acid sequences are aligned and numbered in accordance with the IMGT unique numbering (7) and gaps thereby introduced are indicated by dashes. Amino acids that are part of the CDR1-IMGT (7) are underlined. Dots indicate identity with the AE11F sequence. Canonical structures were determined by automatic canonical structure classification (27).

| scFv clone | Amino acid sequence | Germline origin | Canonical structure |
|------------|--|-----------------|---------------------|
| | 2 3 33334 4 | | |
| | 2 0 45690 4 | | |
| AE11F | SCAASGFI <u>F</u> SEYD--MHWVRQ | IGHV3-30 | 1 |
| E3 | ...V...GSI.SGGYYWS.I.. | IGHV4-31 | 3 |
| E6 | ...V...YSI.SGYY-WG.I.. | IGHV4-b | 2 |
| E10 | ...V...GSI.S.Y--WS.I.. | IGHV4-59 | 1 |
| E11 | ...V...GSI.G.H--WS.I.. | IGHV4-34 | 1 |
| E14 | ...V...GSI.SGGYSWS.I.. | IGHV4-30-2 | 3 |

the grafting as these are only distantly related to the IGHV3 CDR and therefore allow for a higher degree of variability. In addition, genes from the IGHV4 subgroup encode loops of different lengths than genes from the IGHV3 subgroup, including loops of the same length as the ones created by the single-codon insertions in CDRH1, thus enabling a comparison with these modifications. Sequencing of randomly picked clones showed that seemingly functional, *i.e.* in-frame and without stop codons, IGHV3 genes carrying IGHV4-derived CDRH1 sequences were obtained (Table II). However, when analyzing crude expression supernatants of the constructs, it was found that all of the clones had lost the original antigen specificity and instead acquired a polyreactive character (Fig. 3).

To further investigate this polyreactive nature of the CDRH1-grafted clones, two of them, E3 and E6, were produced at a larger scale and purified as monomers to enable structural characterization. These two clones were chosen based on the presence of loop lengths different from the one used by the parent antibody (Table II). As judged by analytical gel filtration, these clones also gave rise to proteins that behaved as scFv monomers (data not shown). The overall secondary structure was determined by CD spectroscopy and was compared with the results obtained with other monomeric scFv. As shown in Fig. 4, the spectra of both of the CDRH1-grafted clones displayed a strong negative signal near 200 nm, which is indicative of unordered polypeptides (33). For a comparison, the spectra of both the parent scFv and the FITC8 scFv displayed a weak negative signal near 217 nm, which is characteristic of the β -sheet conformation of antibody domains (Fig. 4). The same result was also obtained with clones carrying single-codon modifications, such as the AE11F/3-20L1 and the FSV43, which gave rise to nearly identical spectra as the parent scFv (data not shown). When analyzed by DSC, no unfolding temperatures could be determined for either of the E3 or E6 scFv, suggesting that the proteins already were in an, at least partly, unfolded state. Thus, by inserting these only distantly related CDR sequences into the IGHV3 framework, the boundaries that define a stable immunoglobulin fold had apparently been exceeded.

DISCUSSION

Insertions and deletions of nucleotides have recently been shown to be an additional mechanism whereby immunoglobulin V region genes are evolved (8–11) and which may expand the available repertoire of antibody hypervariable loop lengths and structures. Although sequence modifications of this kind, especially insertions, have also been exploited in antibody engineering, knowledge about the effects of these modifications on protein stability and antigen recognition is still limited. Such factors are critical as they determine the success of this mode of molecular evolution, whether employed by nature or by

the molecular engineer. To study the functional consequences of both insertions and deletions in the CDR of human antibodies, we have here made single-codon insertions and deletions as well as more extensive modifications in the CDR of two antibody fragments with different specificities and assessed the thermal stability and the antigen binding properties of the resulting proteins.

The single-codon modifications were well tolerated by the two scFv frameworks as determined by the thermal stability measurements and the high ratio of functional clones despite the fact that they created both loop lengths that do not occur normally within the human IGHV3 subgroup and combinations of loop lengths that do not exist in the human germline repertoire. Insertion of one residue in CDRH2 of the two scFv studied here creates a loop length (CDR2-IMGT length 9 amino acids) that is not naturally encoded by any IGHV genes except for the only member of the IGHV6 subgroup (7). This loop length has been predicted to have its own distinct conformation (canonical structure 5, Ref. 31), but as no immunoglobulin encoded by this gene has been structurally determined, this canonical structure has not been defined. The insertion of one residue in CDRH1 produces a loop length (CDR1-IMGT length 9 amino acids) that occurs naturally within the human IGHV4, but not the IGHV3 subgroup, and which could correspond to canonical structure 2 as judged by the automatic canonical structure classification. This coexistence of canonical structure 2 in CDRH1 with canonical structure 3 in CDRH2 (Table I) does not occur naturally within the human IGHV germline repertoire, although it has been observed in hypermutated antibodies with insertions in CDRH1 (8). In addition, the structure classification also revealed that a large number of the key residue requirements for canonical structure 2 were not fulfilled (27), *i.e.* the thus modified CDRH1 loops either take on structures not covered by the described canonical structures or adopt the observed structure corresponding to this loop length despite the presence of a large number of disallowed amino acids at key residue positions. Irrespective of the circumstances, the insertions in CDRH1 seem to, like the rest of the single-codon modifications, give rise to scFv that are correctly folded and stable.

The fact that the loop lengths that were created by the single-codon insertions are not part of the IGHV3-encoded repertoire does not mean that they are completely unnatural in the context of an IGHV3 framework. Apparently functional antibodies belonging to the IGHV3 subgroup with insertions in CDRH1 and CDRH2 leading to CDR-IMGT loop lengths of 9 amino acids have in fact been described by others (8, 34, 35). As the deletions at position 58 in CDRH2 of both scFv give rise to loop lengths that are used by other members of the IGHV3 subgroup, it is not entirely unexpected that these modifications

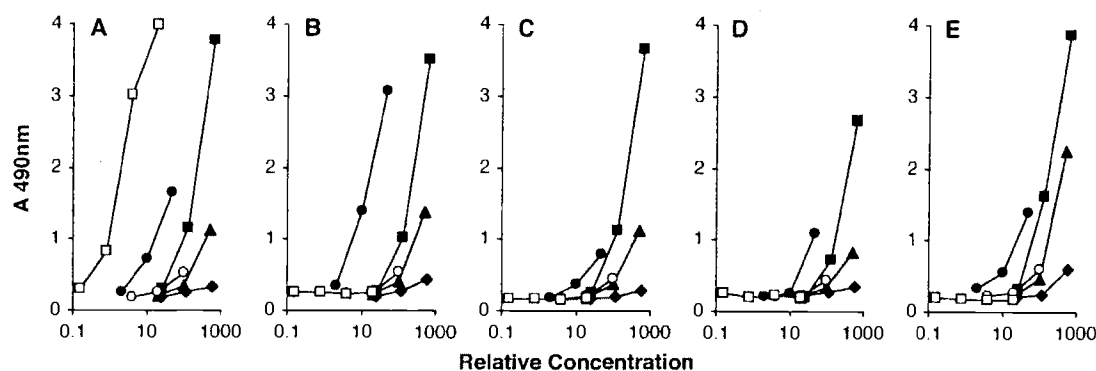


FIG. 3. Clones carrying CDRH1 sequences from distantly related IGHV genes displayed a polyreactive character. Reactivity of the AE11F (□), E3 (■), E6 (○), E10 (▲), E11 (◆), and E14 (●) scFv with streptavidin-bound viral peptide (A), streptavidin (B), BSA (C), ovalbumin (D), and uncoated polystyrene wells (E), as determined by ELISA. Relative concentrations of the expression supernatants were estimated by immunoblotting. The coefficient of variation was below 10% for the whole data set.

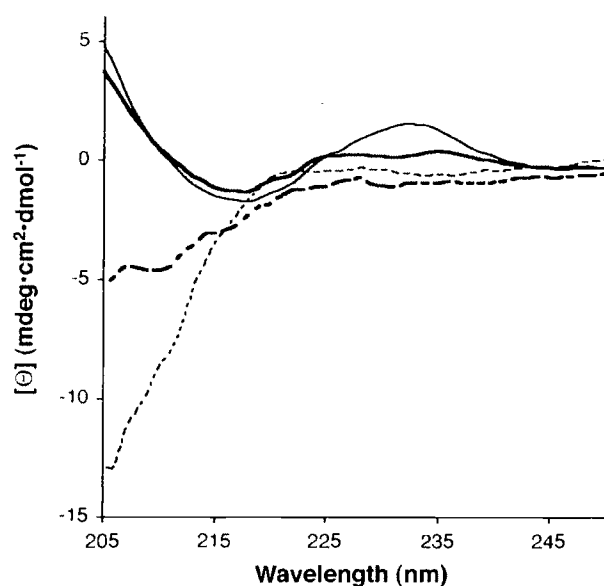


FIG. 4. CD spectroscopy indicated an unordered folding of the CDRH1-grafted clones. CD spectra of purified monomers of the AE11F (thick solid line), FITC8 (thin solid line), E3 (thick broken line), and E6 (thin broken line) scFv in 50 mM sodium phosphate, pH 7.4.

are tolerated by the scFv frameworks studied here. Furthermore, in a previous study, we have found that single-codon deletions, some of which have also been shown to be functional, occur in antibodies belonging to the IGHV3 subgroup at or immediately adjacent to position 58 (12). The single-codon modifications of antibody sequence space we have presented here are in other words highly representative of changes that may occur naturally as a consequence of the somatic hypermutation process.

As some of the single-codon insertions produced loop lengths found in antibodies belonging to the IGHV4 subgroup, we decided to investigate the possibility of using CDRH1 sequences originating from this subgroup to diversify the AE11F scFv. This approach resembles evolution through receptor revision, which occurs *in vivo* (36, 37) and has also been shown to provide a selection advantage *in vitro* (38). However, grafting of CDRH1 loops of different lengths from the IGHV4 subgroup into the IGHV3 framework used by the AE11F scFv resulted not only in a loss of the original antigen specificity but also in the acquisition of a polyreactive character, even when not having been put through a potentially denaturing purification process (39), by the thus modified scFv clones (Fig. 3). This

polyreactivity is most likely due to a destabilized or inappropriately folded V domain, as demonstrated by the CD spectra of two of the clones (Fig. 4). Destabilizing effects of loop grafting into an antibody framework have been reported previously (40), but in that particular case, the grafted sequences were totally unrelated to antibody hypervariable loops. The use of naturally occurring CDR sequences for grafting into immunoglobulin frameworks often ensures that the inserted loops are optimally functional as they have been proofread and selected for functionality during the formation of the B cell receptors. Our data show, however, that the functionality of the grafted loops also depends on the framework they are inserted into even if they are natural immunoglobulin sequences. The reason for the observed effects probably lies in the differences in certain key residues between the IGHV3 and IGHV4 frameworks. In fact, many of the amino acids that differ between the original AE11F sequence and the grafted sequences are residues that are used to define the canonical structures (27, 31). In addition, Tramontano *et al.* (41) have shown that framework residue 80 of the heavy V domain packs against residues in both CDRH1 (position 30) and CDRH2 (position 58) and that it is an important determinant of the conformation of the CDRH2 loop. A subsequent mutational study has also shown that the nature of this residue determines the binding characteristics of an antibody by influencing the conformation of the heavy chain CDR loops (42). The AE11F framework has, like all unmutated antibodies belonging to the IGHV3 subgroup, an Arg at position 80, whereas all genes belonging to the IGHV4 subgroup, from which the CDRH1 sequences were obtained, encode a Val residue at this position in their germline configurations. The larger, charged Arg possibly causes clashes with the IGHV4-derived residues in and adjacent to CDRH1, which leads to an improper fold and poor stability of the resulting scFv product.

In conclusion, we demonstrate here that single amino acid insertions in both CDRH1 and H2 and deletions in CDRH2, which are highly representative of modifications that occur naturally in regions of the hypervariable loops known to be involved in antigen contact (15) during the maturation of B cell receptors, are well tolerated and permit production of stably folded proteins. This is true despite the fact that the thus modified loops do not fulfill the key residue requirements for canonical loops of the corresponding length or are of a length not associated with a known canonical structure (27). This demonstrates the plasticity of antibody V domain frameworks belonging to the important IGHV3 subgroup, which makes up a large fraction of all human antibodies (43), and its capacity to tolerate modifications that expand sequence and structure space beyond the limits set by the germline-encoded diversity.

Based on the similarities with naturally occurring alterations of loop lengths, our results with insertions and deletions in CDRH1, H2, and L1 of the antibody fragments used in this study, and work on an unrelated scFv with a three-amino acid insertion at the beginning of CDRH1 (10),³ our conclusion is that both insertions and deletions can be efficiently utilized in antibody engineering to expand the structural space available to human antibodies as long as attention is paid to key residues in the framework (41). As demonstrated by previous studies on murine antibodies, this approach can be used for improving already existing specificities (17, 44). However, analogously with the correlation between CDR loop lengths and the antigen recognized (16), it is conceivable that it may also be utilized for the construction of antibody libraries specific for a particular class of antigens such as haptens, peptides, or large molecules. Finally, we hypothesize that introduction of novel loop lengths and combinations of loop lengths not encoded by the germline repertoire may also enable the targeting of poorly immunogenic or previously unrecognized antigens and epitopes as entirely new regions of antibody structure space are explored by this mode of sequence diversification.

Acknowledgments—We thank Ola Jakobsson and Micael Owald for technical assistance.

REFERENCES

- Amzel, L. M., and Poljak, R. J. (1979) *Annu. Rev. Biochem.* **48**, 961–997
- Padlan, E. A. (1994) *Mol. Immunol.* **31**, 169–217
- Al-Lazikani, B., Lesk, A. M., and Chothia, C. (1997) *J. Mol. Biol.* **273**, 927–948
- Tonegawa, S. (1983) *Nature* **302**, 575–581
- Berek, C., and Milstein, C. (1987) *Immunol. Rev.* **96**, 23–41
- Nemazee, D., and Weigert, M. (2000) *J. Exp. Med.* **191**, 1813–1817
- Lefranc, M. P. (2001) *Nucleic Acids Res.* **29**, 207–209
- de Wildt, R. M., van Venrooij, W. J., Winter, G., Hoet, R. M., and Tomlinson, I. M. (1999) *J. Mol. Biol.* **294**, 701–710
- Goossens, T., Klein, U., and Küppers, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2463–2468
- Ohlin, M., and Borrebaeck, C. A. K. (1998) *Mol. Immunol.* **35**, 233–238
- Wilson, P. C., de Bouteiller, O., Liu, Y. J., Potter, K., Banchereau, J., Capra, J. D., and Pascual, V. (1998) *J. Exp. Med.* **187**, 59–70
- Lantto, J., and Ohlin, M. (2002) *J. Mol. Evol.* **54**, 346–353
- Wilson, I. A., and Stanfield, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 113–118
- Webster, D. M., Henry, A. H., and Rees, A. R. (1994) *Curr. Opin. Struct. Biol.* **4**, 123–129
- MacCallum, R. M., Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **262**, 732–745
- Vargas-Madrado, E., Lara-Ochoa, F., and Almagro, J. C. (1995) *J. Mol. Biol.* **254**, 497–504
- Lamminmäki, U., Pauperio, S., Westerlund-Karlsson, A., Karvinen, J., Virtanen, P. L., Lövgren, T., and Saviranta, P. (1999) *J. Mol. Biol.* **291**, 589–602
- Ohlin, M., Sundqvist, V. A., Mach, M., Wahren, B., and Borrebaeck, C. A. K. (1993) *J. Virol.* **67**, 703–710
- Ohlin, M., Owman, H., Mach, M., and Borrebaeck, C. A. K. (1996) *Mol. Immunol.* **33**, 47–56
- Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Åberg, A. M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., Danielsson, L., Carlsson, R., and Borrebaeck, C. A. K. (2000) *Nat. Biotechnol.* **18**, 852–856
- Lantto, J., Lindroth, Y., and Ohlin, M. (2002) *Eur. J. Immunol.* **32**, 1659–1669
- Jirholt, P., Ohlin, M., Borrebaeck, C. A. K., and Söderlind, E. (1998) *Gene (Amst.)* **215**, 471–476
- Lantto, J., Jirholt, P., Barrios, Y., and Ohlin, M. (2002) *Methods Mol. Biol.* **178**, 303–316
- Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L., and Conlon, P. J. (1988) *Bio/Technology* **6**, 1204–1210
- McGrew, J. T., Leiske, D., Dell, B., Klinke, R., Krasts, D., Wee, S. F., Abbott, N., Armitage, R., and Harrington, K. (1997) *Gene (Amst.)* **187**, 193–200
- Jirholt, P., Strandberg, L., Jansson, B., Krambovitis, E., Söderlind, E., Borrebaeck, C. A. K., Carlsson, R., Danielsson, L., and Ohlin, M. (2001) *Protein Eng.* **14**, 67–74
- Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **263**, 800–815
- Lantto, J., Fletcher, J. M., and Ohlin, M. (2002) *J. Gen. Virol.* **83**, 2001–2005
- Whitelegg, N. R., and Rees, A. R. (2000) *Protein Eng.* **13**, 819–824
- Tomlinson, I. M., Cox, J. P., Gherardi, E., Lesk, A. M., and Chothia, C. (1995) *EMBO J.* **14**, 4628–4638
- Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewellyn, M. B., and Winter, G. (1992) *J. Mol. Biol.* **227**, 799–817
- Lantto, J., Fletcher, J. M., and Ohlin, M. (2002) *Virology*, in press
- Woody, R. W. (1995) *Methods Enzymol.* **246**, 34–71
- Brezinschek, H. P., Foster, S. J., Brezinschek, R. I., Dorner, T., Domiati-Saad, R., and Lipsky, P. E. (1997) *J. Clin. Invest.* **99**, 2488–2501
- Noppe, S. M., Heirman, C., Bakkus, M. H., Brissinck, J., Schots, R., and Thielemans, K. (1999) *Br. J. Haematol.* **107**, 625–640
- Wilson, P. C., Wilson, K., Liu, Y. J., Banchereau, J., Pascual, V., and Capra, J. D. (2000) *J. Exp. Med.* **191**, 1881–1894
- Itoh, K., Meffre, E., Albesiano, E., Farber, A., Dines, D., Stein, P., Asnis, S. E., Furie, R. A., Jain, R. I., and Chiorazzi, N. (2000) *J. Exp. Med.* **192**, 1151–1164
- Ellmark, P., Esteban, O., Furebring, C., Malmberg Hager, A.-C., and Ohlin, M. (2002) *Mol. Immunol.* **39**, 349–356
- McMahon, M. J., and O'Kennedy, R. (2000) *J. Immunol. Methods* **241**, 1–10
- Helms, L. R., and Wetzel, R. (1995) *Protein Sci.* **4**, 2073–2081
- Tramontano, A., Chothia, C., and Lesk, A. M. (1990) *J. Mol. Biol.* **215**, 175–182
- Xiang, J., Sha, Y., Jia, Z., Prasad, L., and Delbaere, L. T. (1995) *J. Mol. Biol.* **253**, 385–390
- de Wildt, R. M., Hoet, R. M., van Venrooij, W. J., Tomlinson, I. M., and Winter, G. (1999) *J. Mol. Biol.* **285**, 895–901
- Parhami-Seren, B., Viswanathan, M., and Margolies, M. N. (2002) *J. Immunol. Methods* **259**, 43–53

³ J. Lantto and M. Ohlin, unpublished work.